THE USE OF ORAL RETELLINGS OF BASAL STORIES AS AN INFORMAL READING DIAGNOSTIC TECHNIQUE

By

ROBERT GERARD CARROLL

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1985

Copyright 1985

by
Robert Gerard Carroll

To Gennette

ACKNOWLEDGMENTS

The author of this investigation is a very fortunate individual. More people have contributed more assistance and support to him during the last three years than can possibly be acknowledged here. Nonetheless, there are a few people who deserve both recognition and a permanent debt of gratitude.

Dr. H. Thompson Fillmer has served as the chairman of the author's doctoral committee for three years. During that time, he freely gave of his time, expertise, insight, patience and friendship while demanding excellence in return. Without Dr. Fillmer's guidance, this project and this degree would not have been completed.

To a great extent, much of the intellectual rigor displayed in this document can be credited to Dr. Nora Hoover. Throughout the past three years she continually coaxed, challenged, and inspired the author by her vigorous pursuit of excellence in teaching and research and by her refusal to accept anything but the author's best work.

Dr. Stephen Olejnik had incredible demands placed on his time and yet he was willing to give it freely and in abundance when the author needed help. His knowledge and experience saved the author from numerous mistakes and the efficiency with which this project functioned is to his credit.

Dr. Suzanne Kinzer has given the author much needed advice, guidance, and friendship. She has also been an example of what a professional educator should be.

Abundant thanks and appreciation are owed to Dr. David C. Smith, Dean of the College of Education. The author served Dean Smith as his graduate assistant for the past three years. During that time, the author has learned much about integrity, honesty, courage during adversity, and unwavering dedication to excellence. The author arrived in Gainesville thinking of himself as a professional in reading; he now sees himself as a teacher educator who specializes in reading. That change, though subtle, has already had a profound impact on the author's career aspirations and is to be credited to the influence and wisdom of Dean Smith.

Dr. Arthur J. Lewis is the most inspiring individual the author has ever met. He is a researcher, a theorist, a philosopher, a futurist, a teacher, and a marvelously kind and gentle man. For some reason, he befriended the author and has been a confidant, an advisor, and an inspiration from the first.

Special thanks are due to a number of persons outside of the University of Florida. Among them are

Dr. Raymond L. Kimble and Dr. Joseph A. Fusaro of the Department of Education at the University of Scranton for

introducing the author to the teaching of reading and for convincing him that he really could get a doctoral degree;

Annette and Vincent Carroll, the author's parents, for providing an endless supply of love and warmth and for supporting the author when he needed them most;

Gennette Gailey, the author's wife, for showing him how good and how deep love can be and for keeping him afloat during three sometimes turbulent years;

Paul Pawlowski, Maryclare Noone, and Joseph Hogan, three of the finest people alive, for being themselves, for supporting the author during the most trying years of his life, and for continuing to do so today; and

Rebecca Enneis, for helping in the preparation of this manuscript, but more importantly for being such a wonderful person in such a beautiful way.

To all of these magnificent people the author readily admits that whatever he may ever be able to accomplish in the future will represent his continuing attempts to give to others what so many have given to him.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....iv

ABSTRACT.....x

CHAPTER PAGE

ONE INTRODUCTION AND JUSTIFICATION.....1

Introduction.....1
The Problem.....2
Statement of the Problem.....2
Hypotheses.....2
The first hypothesis.....2
The second hypothesis.....3
The third hypothesis.....3
The fourth hypothesis.....4
Significance of the Problem.....5
Delimitations, Limitations, Assumptions, and Definitions.....14
Delimitations.....14
Limitations.....16
Assumptions.....17
Definition of Terms.....18
Summary.....19

TWO REVIEW OF THE LITERATURE.....22

Introduction.....22
Part I: Schema Theory and Story Structure.....23
Part II: Using Retellings in the Classroom.....60
Part III: Critical Variables in the Assessment of Retellings.....78
Sequence.....79
Inferences.....81
Intrusions.....83
Metacognition.....85
Vocabulary.....87
Background Knowledge.....89
Summary.....91
Summary of Chapter Two.....92

THREE PROCEDURES.....94

Introduction.....94
Development of the Checklist.....95

The Training Sessions....97
The Control Group....98
The Experimental Group....98
Experiment 1....100
Sample....101
Materials....101
Instrumentation....101
Methodology....101
Data Analysis....103
Experiment 2....103
Sample....103
Materials....104
Instrumentation....104
Methodology....104
Data Analysis....105
Experiment 3....105
Sample....105
Materials....106
Instrumentation....107
Design of the Experiment....107
Methodology....108
Data Analysis....112
Experiment 4....113
Sample....113
Materials....114
Instrumentation....114
Design of the Experiment....114
Methodology....115
Data Analysis....116
Summary....116

FOUR RESULTS OF THE INVESTIGATION....118

Introduction....118
Experiment 1....119
Experiment 2....124
Experiment 3....124
Experiment 4....128
Qualitative Data....129

FIVE SUMMARY, IMPLICATIONS, CONCLUSIONS, AND RECOMMENDATIONS....136

Introduction....136
Summary of the Investigation....136
The Research Question and the Four Hypotheses....136
The Theoretical Foundation of the Investigation....139
Procedures and Findings....144
Implications....147

Conclusions and Recommendations.....157
Conclusions.....157
Recommendations.....161
Summary.....163

APPENDICES

A CARROLL ORAL RETELLING CHECKLIST.....166

B STRUCTURED INTERVIEW INSTRUMENT.....167

REFERENCES.....169

BIOGRAPHICAL SKETCH.....177

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

THE USE OF ORAL RETELLINGS OF BASAL STORIES AS AN INFORMAL READING DIAGNOSTIC TECHNIQUE

By

Robert Gerard Carroll

August 1985

Chairman: Dr. H. Thompson Fillmer

Major Department: Instruction and Curriculum

This project was designed to answer the following question: Can the analysis of students' oral retellings of basal stories be used by teachers as an effective informal reading comprehension diagnostic technique? The Carroll Oral Retelling Checklist (CORC) which is based on story-schema research was designed and used in the investigation. The first experiment correlated fifth graders' CORC scores with standardized comprehension scores. Two correlations were computed, one with all 30 data points ($r = .57$) and one with three apparently invalid scores removed ($r = .78$). These significant results supported the construct validity of the CORC. The second experiment examined the inter-rater reliability of the CORC when five raters independently assessed 10 retellings of a story. All correlations were significant with a range of .86 to .98.

The third experiment involved 30 fourth- and fifth-grade students who were below average in reading. Six teachers were given a brief training session and then each one used the CORC for 12 weeks with four to six randomly selected students. Subjects were asked to read a basal story and to retell it to the teacher who tape-recorded the session. Teachers analyzed the retellings and planned remediation using the CORC. A posttest retelling compared the experimental group with 30 controls; an ANOVA revealed a significant difference existed in favor of the treatment group. Quantitative and qualitative data suggested that teachers had successfully used the CORC to diagnose comprehension and plan remediation. All teachers adopted a story-structure approach to comprehension instruction and attributed student growth to improvements in students' story schema.

A fourth study compared written diagnoses of experimental- and control-group teachers to test for increased diagnostic insights due to use of the CORC. Results, however, were inconclusive and some questions were raised concerning the ability of the posttest to provide an assessment of teachers' diagnostic abilities.

Overall, the CORC proved to be a valid and reliable instrument that can be used to diagnose comprehension weaknesses and prescribe remediation. Support was also

found for the positive effects on comprehension of instruction in story-structure for below average readers.

CHAPTER ONE
INTRODUCTION AND JUSTIFICATION

Introduction

Educational assessment, and reading assessment in particular, is used to inform three kinds of decisions: administrative, diagnostic, and selection/classification (Johnston, 1983). The first and third of these typically involve the use of standardized tests which aid school systems in making large scale decisions. Diagnostic decisions, however, are ipsative in nature and, while standardized procedures can certainly be of help, informal techniques can yield more highly individualized information (Pikulski & Shanahan, 1982).

One method for the informal assessment of reading comprehension that has drawn recent attention is that of analyzing students' oral retellings of stories. However, while several authors have offered the approach as an effective diagnostic technique (Clark, 1982; Irwin & Mitchell, 1983; Marshall, 1983), there has been little statistical evidence offered in support of using oral retellings as a source of diagnostic data. It was the purpose of this research project to investigate the analysis of elementary children's oral reports of basal stories as a classroom diagnostic procedure.

The Problem

Statement of the Problem

This research project was designed to investigate the following question: Can the analysis of students' oral retellings of basal stories be used by classroom teachers as an effective informal technique for diagnosing reading comprehension? An examination of this question required (a) developing an instrument for use in the assessment of retellings, (b) establishing the reliability and validity of the instrument, (c) training a number of educators in the use of the instrument, (d) assessing the impact of the technique on students' oral retelling behaviors, and (e) examining the technique's influence on teachers' informal diagnostic insights.

Hypotheses

The primary question of this research project gave rise to at least four additional questions which constituted the specific foci of this study. Each of these questions implied a separate hypothesis and a corresponding investigation. For ease of analysis and interpretation, each was stated as a null hypothesis and all were tested at the .05 level of significance.

The first hypothesis

The first question of interest in this research project was stated as follows: Does the oral retelling checklist developed for use in this project have the capacity to

differentiate between good and poor reading comprehension as measured by a standardized test? That is, how valid is the instrument as compared to a standardized test? The first hypothesis, then, was

Hypothesis 1. There will be no statistically significant correlation between subjects' standardized reading comprehension test scores and scores assigned to them through the use of the oral retelling technique.

The second hypothesis

The first hypothesis relates to the validity of the instrument in question. A logical second step was to consider the reliability of the instrument. The second question addressed by this project, therefore, was as follows: To what degree would a group of educators, trained in the use of the oral retelling technique, agree on the relative quality of comprehension revealed by a series of students retelling a single story? That is, what level of inter-rater reliability is associated with this instrument? The corresponding hypothesis was

Hypothesis 2. There will be no statistically significant correlation among the scores assigned to 10 different retellings of a single story by five trained users of the oral retelling checklist.

The third hypothesis

After the validity and reliability of the instrument were established, the focus of this investigation shifted to

its practical applications in the classroom. The first area of interest in this regard was the impact of the use of the technique on students' oral retelling behaviors. At no time during this investigation were students instructed in how to retell a story or in the components of a retelling or a story grammar. Rather, teachers were encouraged to use the retellings to diagnose students' reading comprehension strengths and weaknesses, and to use that diagnosis to plan subsequent instruction whenever possible. The question, then, was as follows: Does repeated exposure to the oral retelling diagnostic technique lead to more complete retellings by students? This led to the third hypothesis.

Hypothesis 3. There will be no statistically significant difference between the quality of the final retellings of students who have been repeatedly exposed to the oral retelling technique and retellings of the same story done by students who have not been exposed to the technique.

The fourth hypothesis

Aside from the effects of the procedure on students' oral retelling behaviors, an additional concern was its effects on teachers' diagnostic insights. Therefore, the fourth question addressed in this project was as follows: Does the repeated use of the oral retelling diagnostic technique improve teachers' ability to make informal

assessments of students' reading comprehension strengths and weaknesses? The fourth hypothsis, then, was

Hypothesis 4. There will be no statistically significant difference between the quality of the diagnostic assessments developed by teachers who have used the oral retelling technique and by those who have not, as judged by a panel of experts in reading.

Significance of the Problem

In order to establish the significance of this investigation, it is necessary to place it in the context of prior theory and research. In Chapter Two, that theory and research is reviewed at length. However, the following is a brief review of a few key studies that provided the foundation for this study. The review begins with a brief examination of schema theory. Next, research concerning one particular schema, a schema for stories, is reviewed and some of the effects of that schema on readers' comprehension and recall are outlined. Finally, the use of retellings as an approach to comprehension assessment is reviewed with particular attention to its use in the classroom.

The theoretical foundation for this investigation begins with Bartlett (1932) and his seminal work in schema theory. A schema is thought to be a psychological blueprint for a concept or situation. When encountering some stimuli, an individual is believed to search all previously held schemata until one is found that fits the incoming data.

According to Bartlett (1932), that schema acts as a framework or set of expectations for relevant information. For example, a "purchase" immediately brings to mind a seller, a buyer, and some merchandise. Bartlett believed that schemata are not static constructs. Instead, he believed that they are active and evolving, providing organization for incoming data and changing to accommodate unique experiences. Modern treatments of schema theory by researchers interested in cognition and reading have expanded the theory (Adams & Collins, 1979; Anderson & Pearson, 1984; Rumelhart & Ortony, 1977) and provided empirical evidence of its effects on comprehension (Kintsch, 1977; Kintsch & Greene, 1978; Mandler, 1978; Rumelhart, 1975, 1980).

One particular schema which is believed to have critical implications for reading is that of a schema for stories (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979). Several scholars have contended that readers possess a schema for stories. That is, readers have a set of expectations about the structure of narratives which provides a framework for processing and classifying data, retaining it, and reproducing it in a coherent fashion. Mandler and Johnson (1977), for example, found that while adult readers retain more than children, the patterns of their recall (i.e., the saliency of the individual elements

of a story's structure) were almost identical for all age groups studied.

In a subsequent study, Mandler (1978) replicated these results and expanded upon them. In one portion of the study, she presented readers with a story that had the events of two different episodes interwoven within each other. During recall, subjects tended to reproduce the story as if the two episodes had been presented separately. She concluded that this reordering of events into a more coherent structure was the result of the influence of the reader's story schema.

Stein and Glenn (1979) found further support for the effects of a story schema upon recall. In the first of two studies, they asked children to listen to a story and then orally recall it, trying to reproduce the original as accurately as possible. Then the exercise was repeated with a second story. One week later, the children were asked to recall the two stories again. Their analyses showed that recall followed the same general pattern of saliency as reported by Mandler and Johnson (1977). They also found that the saliency of individual statements made by children over time was quite high. Stein and Glenn (1979), however, moved beyond what children recalled to an analysis of what additions children made during recall. They found that the Internal Reactions category, which had been the least well recalled in their own study of category saliency, as well as

in Mandler and Johnson (1977), was the most frequently added category. This finding suggested that children were more aware of internal responses than had been previously believed. However, instead of explicitly recalling the internal responses in the story, children seemed to be supplying responses of thir own.

In their second study (Stein & Glenn, 1979), the authors were concerned with what children believed to be most important in a story and children's understanding of the causal relationships in a story. Subjects listened to four stories, and after hearing each one were asked to name the most important thing in the story. Then they were asked for the second and third most important things. After they heard all of the stories they were asked a series of probes concerning causal relationships. The results showed that students placed a high degree of importance on internal responses, confirming their conclusions from the first study, and that students seemed to be fully aware of the causal links in a story.

The Stein and Glenn (1979) study is a complex and an important one and it is reviewed in much greater depth in Chapter Two. However, from the results reported here it is evident that the authors found strong support for the existence of a schema for stories and for the effects of that schema on comprehension, retention, and recall.

In summation, schema theory seems to provide a robust explanation for a number of phenomena in reading comprehension. Of particular import to this investigation is the existence of a schema for stories. There is some evidence that readers' sense of the structure of stories has a powerful impact on individuals' expectations while reading stories and on their processing, retention, and reproduction of the content of stories.

Many of the researchers who have investigated schema theory have used retellings, either oral or written, as a research technique (Bartlett, 1932; Mandler & Johnson, 1977; Stein & Glenn, 1979). While there has been increasing interest in recent years in the use of retellings both for assessment and for instruction in the classroom, the use of retellings has a long history.

An early explanation of the use of retellings as a diagnostic device appeared in an article by Courtis (1914). He recommended a four step procedure that included (a) counting the number of "idea units" found in the text to be read, (b) having a child read the text, (c) asking the child to write down everything that could be remembered from the passage, and (d) determining the number of idea units accurately reproduced by the student. According to the author, this would yield a measure of the accuracy of the child's reading comprehension.

H. A. Brown (1914) reported a similar procedure. He asserted that reading efficiency can best be determined by considering three factors: rate of reading, quantity of reproduction, and quality of reproduction. Accordingly, a child would be told to read from a passage for one minute and then to write down everything that could be remembered. Next, the teacher would examine the written recall and determine the number of idea units written by the child as compared to the total number of idea units in the passage. Finally, each idea unit in the recall would be compared to the original. Only those ideas which were completely accurate and included all original material were counted. The results of the former comparison were intended to produce a "quantity measure," while the latter produced a "quality measure" of comprehension.

A third approach to diagnosis that employed written free recalls was detailed by Starch (1915). His method involved determining the number of words in the recall that accurately reproduced the original without adding new ideas or repeating ideas previously recorded. The total number of "correct" words was used as an index of comprehension. The author was not specific as to how such a raw score would provide an estimate of comprehension, but presumably the teacher would compare this number to the total number of words in the passage to determine a proportion of words recalled. A problem with this technique is that it is

possible to get a proportion in excess of 1.0 by writing, for example, "The small black animal with the big white stripe sprayed me," for "The skunk sprayed me." The resulting "proportion" of 2.75 is nonsensical.

Interestingly, near the end of this article Starch noted that written recalls may not be the best method of tapping student comprehension. He wrote: "Ideally, the comprehension should perhaps be tested by having each pupil state orally in his own words what he had read and by having a stenographic report of his statements" (Starch, 1915, p. 22). This alternative is rejected, however, as being impractical. Still, it points out that at least one diagnostician believed that oral recalls may more accurately reflect reading comprehension than written ones.

This foreshadowed an objection which was raised by Kelly (1916) and continues to be an issue today (Johnston, 1984). It concerns the production skills required to produce a written recall. Kelly contended that written recalls require production skills which can hinder an accurate diagnosis by inhibiting the student from making a full report of everything that was comprehended during reading. He argued that comprehension and reproduction were very different skills and that the diagnosis of one should require as little as possible of the other. He proposed using tests that employed a multiple choice format. Unfortunately, many of the sample items included by the

author required little more than the ability to recognize a correct answer from among three choices. Such items may reflect recognition rather than understanding. However, because of objections such as those noted by Kelly (1916) the use of retellings to assess comprehension nearly disappeared from the literature until the early 1970's.

The recent interest in students' free recall protocols as a research and diagnostic device was kindled by Y. Goodman and Burke (1970). They included oral free recalls of text, now referred to as retellings, in their Reading Miscue Inventory as a method of diagnosing reading comprehension. Since that time, both Kenneth and Yetta Goodman have written extensively about the technique (e.g., K. Goodman & Y. Goodman, 1977; Y. Goodman, 1982). They urged both teachers and researchers to become familiar with and to use retellings as a measure of comprehension.

An early study of the use of retellings in the classroom was reported by Zimiles and Kuhns (1976). They worked with 576 children from six to eight years of age. Among their many findings was that retelling stories immediately after hearing them significantly improved children's recall. Results also showed that seven-year-olds who retold stories retained significantly more than eight-year-olds who did not retell them. Morrow (1985a, 1985b, 1985c) reported two studies in which retellings were used. In both studies the treatment group retold stories

immediately after hearing them while the control group illustrated some aspect of the story. In the first study retellings were free recalls; in the second retellings were cued. The experimental group performed significantly better than the control group on a comprehension posttest in the first study. In the second, a storytelling activity was added as an additional posttest measure. The experimental group included significantly more structural elements, maintained a more correct sequence, and wrote with greater syntactic complexity than did the control group.

A number of other authors have also reported success with retellings (Clark, 1982; Irwin & Mitchell, 1983; Marshall, 1983). Marshall, for example, reported success with a retelling approach that included a checklist and probe questions built to reflect research concerning the structure of stories. Irwin and Mitchell (1983) reported a method of assessing the richness of retellings which could give teachers information on their students' generalizations, understanding of the thesis of the story, and their recall of major events and supporting details. Later, the same authors (Irwin & Mitchell, 1984) developed an assessment instrument for use with retellings that focused, among other variables, on accuracy of recall, inferences, background knowledge, and vocabulary. Clark (1982) had some success with an approach that he called the "Frecall Method." It involved dividing a passage into

"pausal units" (Clark, 1982, p. 436), having students read and retell the passage, and scoring the retelling for inclusion of each unit. The proportion of units recalled served as the measure of comprehension. A quality measure was also calculated by rating each unit for importance and determining the mean importance level of those that were recalled.

In general, retellings have been used in reading research for about 70 years. Recently, tentative evidence has emerged that retellings may also prove to be of some value to classroom teacher as a diagnostic technique. The powerful effects of story structure on comprehension, coupled with the diagnostic potential of retellings, may result in an effective tool for classroom use. The purpose of the current research was to investigate that potential.

Delimitations, Limitations, Assumptions and Definitions

Delimitations

This study was confined to the development, testing, revision, and evaluation of the analysis of oral retellings as an informal procedure for assessing children's understanding of basal stories. The study examined the impact of that procedure on the ability of teachers to listen to a retelling and to make an informal assessment of the comprehension exhibited by the child. Generalizations to other diagnostic abilities would be inappropriate.

Any conclusions that are drawn from this investigation are applicable only to teacher and student populations that approximate those described in Chapter Three. That population included teachers of fourth and fifth grade pupils attending basic skills classes in a medium size county school system in northern Florida. All students in this study were determined to be at or below the 44th percentile in reading as measured by the Metropolitan Achievement Test (Prescott, Balow, Hogan, & Farr, 1978). Generalizations to other student populations may be inappropriate.

Since the purpose of this study was to develop a procedure that can be used during a typical reading lesson, it was restricted to include only stories that are included in the basal readers used in this specific county school system. Furthermore, the checklist used in this study was designed to reflect story grammar research. As such, the reading material was narrative in style. No purely expository, persuasive, or descriptive material was used. Therefore, all results and conclusions reported in this paper are valid only insofar as they relate to narrative material.

Only oral retellings of stories were examined in this study, and then only after the material had been read silently. No conclusions about the reliability, validity, or impact of assessment procedures which involve either oral

reading or written retellings should be drawn from this research.

Limitations

The limitations of this study relate to the nature of the experimental sample, the restricted range of grade levels and ability levels included in the study, and the specific focus of the study itself.

One set of limitations is inherent in the use of students enrolled in basic skills classes. All students scored at or below the 44th percentile in a recognized standardized test of reading skills, thus producing a restricted range of ability levels that does not approximate that which is found in most classrooms. The sample student population included only fourth- and fifth-grade pupils, and no generalizations should be made beyond those grade levels. All students who participated in this study have English as their primary language. Therefore, generalizations to other student populations might be unwarranted.

Another set of limitations arises from the nature of the sample teacher population. First, the sample size was small and generalizations based on such a sample must be viewed as tentative. Second, since participation in this research project was voluntary on the part of the teachers and students, it must be recognized that the potential exists for a selection by treatment interaction that could result in a threat to the external validity of the study.

The focus of this study on oral retellings precludes generalizations to written retellings of basal stories despite some obvious similarities between the two retelling techniques. It is recognized that the production skills used in oral and written retellings are somewhat different and the assessment of the latter would have to include techniques beyond the scope of this study. Furthermore, this study addresses only those facets of comprehension represented on the assessment instrument. All conclusions and generalizations must be limited to them alone.

Finally, it must be stressed that this procedure is an informal diagnostic technique. As such, it cannot be seen as a replacement for more formal approaches to reading assessment. It is recognized that, although this study indicates that there is some value to the technique, it should be viewed as a supplement to other forms of assessment.

Assumptions

Three assumptions, in particular, were critical to this investigation. The first is that oral retellings of stories adequately reflect students' reading comprehension. The second assumption is that students who have adequately comprehended a story will retell it using an acceptable story grammar format. These two assumptions are addressed in Chapter Two.

The third assumption which is fundamental to this investigation is more difficult to substantiate. It is that when teachers gather diagnostic data about their students they will use the information to remediate any weaknesses they identify. The degree to which teachers in this study used the results of the procedure to plan remediation was the focus of the final two analyses reported in this study.

Definition of Terms

Expert. In this study an expert in the field of reading is defined as one who meets at least two of the following three criteria: (a) possesses a doctoral degree in education with a concentration in reading, (b) teaches graduate level courses in reading on a full time basis, (c) has published research on reading and the teaching of reading.

Informal assessment. This investigation uses the term "informal assessment" as it was used by Nancy Marshall. She defined it as "a set of guidelines or procedures that enable the teacher to make decisions about student performance based upon informal observation of student behavior in the classroom" (Marshall, 1984a, p. 80).

Intrusion. An intrusion is a comment or item of information introduced into a retelling that represents the student's opinions, beliefs, attitudes, and/or metacognitive awareness. Intrusions may or may not be relevant to the story being retold.

Metacognition. In this study, metacognition is defined as "the knowledge and control the child has over his or her own thinking and learning activities, including reading" (Baker & Brown, 1984, p. 353). Emphasis must be given to the fact that two separate phenomena are involved: "knowledge about cognition and regulation of cognition" (Baker & Brown, 1984, p. 353).

Retelling. For the purposes of this study retelling refers to the reconstruction of a story either orally or in writing by a child who has read that story silently

Schemata. The term "schemata" is used in this study to refer to abstract knowledge structures possessed by an individual and stored in memory (Anderson & Pearson, 1984).

Story. In connection with the experimental procedures of this study the term "story" refers to narrative passages, either true or fictional, contained in the basals used in the experimental and control classrooms.

Story grammar. The term "story grammar" is defined as "a grammar designed to specify relations among episodes in a story and to formulate rules for generating other stories" (Harris & Hodges, 1981, p. 311).

Summary

Recent years have seen an increase in the use of oral retellings as an instrument for research into reading comprehension. Until this time, most of that research has focused on what oral retellings reveal about reading

comprehension without specifically applying the technique to informal assessment. While many researchers have argued that the analysis of oral retellings can be used successfully by classroom teachers, there has been little statistical evidence offered in support of that stance. This investigation offers some preliminary evidence of the reliability and validity of the technique and examines the impact of its use on teachers' diagnostic insights and the quality of students' retellings.

The first chapter of this document contains an introduction, the statement of the problem, the questions and hypotheses to be considered, background information and a justification for the study, the delimitations and limitations of the study, the definitions of terms, and a summary. The second chapter contains a review of the pertinent literature. The review has three primary foci: (a) schema theory, its manifestation in a schema for stories, and its use by readers during comprehension and recall; (b) the use of retellings in the classroom; and (c) brief summaries of the importance of metacognition, inferences, vocabulary, background knowledge, and an appropriate sequence to the comprehension and reproduction of stories; and a brief summary of the importance of analyzing intrusions made by children during retellings. The third chapter includes the methods and procedures used in the study and the statistical design used to analyze the

data. Chapter Four reports the results of the study, and Chapter Five discusses those results and the conclusions that can be drawn from them. The instrument used in this study appears in Appendix A. A questionnaire used during structured interviews during Experiment 3 appears in Appendix B.

CHAPTER TWO
REVIEW OF THE LITERATURE

Introduction

This research project was initiated to test the assertion made by some recent authors that oral retellings of narrative material can be used as an informal method of diagnosing reading comprehension and guiding instruction. Accordingly, four studies were developed to test various aspects of the technique. Two preliminary studies tested the validity and reliability of an assessment procedure employing a checklist to organize teachers' analyses of oral retellings. A third study was designed to test the effects on students of retelling stories over a 12 week period. It included the collection of both quantitative and qualitative data. The final study tested the effects of repeated use of the technique on teachers' diagnostic insights.

These four studies draw on a body of theory and research relative to the nature of students' comprehension of narrative material. In this chapter, that literature is reviewed under three headings. Part I examines schema theory, its application to the structure of stories, and its impact on comprehension and recall. Part II examines the use of retellings, often in conjunction with story schema, as a diagnostic technique. Part III examines some of the

critical factors involved in the assessment of comprehension through the use of retellings. It should be noted that the last of these parts does not constitute an exhaustive review of the literature on metacognition, background knowledge, vocabulary, etc. Rather, it is meant only to demonstrate the importance of these factors either to the reading comprehension process or to the analysis of oral recall. Therefore, only a few important papers are reviewed in each section.

Part I: Schema Theory and Story Structure

It has been postulated by some that all experiences are organized in the mind in cognitive structures called schemata. These schemata provide frameworks that make sense of new experiences by comparing them to old ones and classifying them into an appropriate slot in memory to be recalled when needed. One such schema, a schema for stories, is thought to help children understand stories by providing a set of expectations that guide the comprehension process. Part I of this review focuses on schema theory in general and on a schema for stories in particular, especially with respect to its impact on reading comprehension. Part I concludes with an examination of some research into the use of story structure in the classroom.

Most reading professionals give credit to Sir Frederic Bartlett (1932) for introducing the term "schema" to modern learning theory and for laying the initial foundation for

schema theory. Bartlett collected data on a large number of adult readers in a series of five experiments. Based on his data he proposed a theory of remembering to account for what he observed.

Bartlett began by defining a schema as "an active organization of past reactions, or of past experiences, which must always be supposed to be operating in any well adapted organic response" (Bartlett, 1932, p.201). He rejected the notion that memory is reduplicative and argued instead that it is actually an act of construction in which "condensation, elaboration, and invention are common features" (Bartlett, 1932, p. 205). He contended that the evolving nature of schemata requires that an "organism has somehow to acquire the capacity to turn round upon its own 'schemata' and to construct them afresh" (Bartlett, 1932, p.206). He added "I wish I knew exactly how it was done" (Bartlett, 1932, p. 206).

Schemata, Bartlett theorized, act in unison with the immediately preceding impulse to make possible a specific adoptive response, producing an orientation for the organism toward whatever is the focus of attention. Mental images play a critical role in creating new schema, in assimilating new information into old schemata, and in recalling items out of schemata. Finally, schemata are "interconnected, organised together and display instinctive tendencies,

interests, and ideals which build them up, an order of predominance among themselves" (Bartlett, 1932, p. 212).

Two of the most frequently cited modern articulations of schema theory, at least as far as reading professionals are concerned, are Rumelhart and Ortony (1977) and Rumelhart (1980). As Rumelhart and Ortony pointed out, many authors in the cognitive sciences, especially those concerned with artificial intelligence, have presented theoretical descriptions of "frames" (Minsky, 1975), "schema" (Bobrow & Norman, 1975), and "scripts" and "plans" (Schank & Abelson, 1975) which are not substantively different from their own schema theory. Therefore, for purposes of this review, the following summary of the nature of schema, its characteristics and variables, is based on the Rumelhart and Ortony (1977) and Rumelhart (1980) papers. Further, because the focal point of this review is on one specific schema, story structure, this description emphasizes those elements of the theory that have the greatest implications for reading comprehension.

Schemata can be defined as "data structures for representing the generic concepts stored in memory" (Rumelhart & Ortony, 1977, p. 101). A schema is comprised of the web of interrelationships that is thought to organize the various elements of the concept. The authors listed four characteristics of schemata:

1. schemata have variables
2. schemata can embed one within the other
3. schemata represent generic concepts which, taken all together, vary in their levels of abstraction; and
4. schemata represent knowledge, rather than definitions. (Rumelhart & Ortony, 1977, p. 101)

To these, Rumelhart added:

5. schemata are active processes; and
6. schemata are recognition devices whose processing is aimed at the evaluation of their goodness of fit to the data being processed. (Rumelhart, 1980, p. 41)

Rumelhart (1980) compared a schema's variables to characters in a play. Just as numerous individuals can play a particular role, so too can a variable be filled or instantiated by a number of possibilities. Take, for example, the basic "roles" of a STRIKE OUT in a baseball game. There must be a PITCHER, a CATCHER, a BATTER, and STRIKES. These roles can be played by a number of actors. A STRIKE can be a "swinging strike, " a "called strike" or a "foul ball," but they all fulfill the basic requirement of a strike. Of course, a STRIKE OUT is one variable of a larger schema, that of a BASEBALL GAME, which itself is a variable within the even larger schema of SPORTING EVENTS. This illustrates the second characteristic; schema can be embedded within each other. Further, the fact that each larger schema represents a greater level of abstraction illustrates the third characteristic listed above.

Rumelhart and Ortony (1977) are careful to point out that schemata are not dictionary entries or lists of

necessary criterial attributes. Rather, they represent a person's knowledge about a certain concept. Hence, if a typically essential attribute is missing one can either infer its existence or adapt to the new situation by altering the former schema to accommodate the current information. While a CATCHER is normally an integral part of a STRIKE OUT it is at least possible, on three consecutive foul bunts, to have a STRIKE OUT without a CATCHER being involved. That does not render the situation incomprehensible, rather it is merely a unique instantiation of a common variable in a BASEBALL GAME.

It is certainly possible that some future readers of the above paragraph may not be very knowledgable about baseball and so not realize that a foul ball on a third-strike-bunt counts as a strike out. If so, those readers will have to reorganize their existing schema to incorporate this unique instantiation. This illustrates the active nature of schemata.

The final characteristic listed by Rumelhart (1980) involves the role of schemata in assessing new information, comparing it to currently held schemata, and determining whether the new stimulus is an instantiation of one of the old schemata. The fan can see the umpire's raised hand, knows "the count," and decides that this is an exemplar of a STRIKE OUT. This function of schemata enables the individual to treat most experiences as being like some

prior experiences rather than as a continual series of entirely unique events. It helps people make sense of a never-ending flow of stimulation.

The above description of a schema is cursory to be sure, but is sufficient for purposes of this research. Theoretical issues concerning the creation of a new schema, for example, or the rewriting of schema theory into some mathematical formulation, are beyond the scope of this review. The critical point here is that reading theorists have widely used the notion of schemata to explain various phenomena within the reading process. One such paper (Adams & Collins, 1979) is often cited as an important articulation of the role of schema theory as it applies to the reading process.

Adams and Collins began by rejecting the notions that the component levels of processing must be organized hierarchically, that the attainment of one level implies the attainment of all lower levels, and that the reverse is not necessarily true. Instead, they argued that processing at each level is aided by both higher and lower levels simultaneously. When one is reading a passage for meanings the process is not strictly from letter to word to meaning. Rather, there are both top-down and bottom-up forces in motion. The former is making hypotheses and seeking answers; the latter is providing data. Only schema theory,

according to the authors, is powerful enough to support the required interactions.

Adams and Collins (1979) expressed their agreement with Norman and Bobrow (1975). Both sets of authors based their argument on three assertions. First, the system, if it is to maintain coherence, must be purposeful. Second, to select among many potential purposes, the system must have some global self-awareness. Third, some mechanism must have access to all schemata in memory in order to guide interpretation. For these reasons, the theory of the cognitive structure as a limited capacity processor is rejected by Adams and Collins (1979) and Norman and Bobrow (1975). Schemata, they assert, must culminate in a central omniscient processor. Such a processor would not be bound by a linear model of reading. It would enable both top-down and bottom-up processing to occur, thus taking advantage of graphic cues and schemata stored in memory simultaneously.

Schema theory fits this vision of an omniscient processor particularly well. The belief, for example, that a schema can be embedded within a larger schema at a greater level of abstraction allows for simultaneous movement among levels of schemata which further allows for both top-down and bottom-up information processing. A reader can obtain information from text, search the memory for an appropriate schema, call one into play, and begin to fill in its slots. The slot-filling activity can be data-driven or resource-

driven (Adams & Collins, 1979). In the former case, slots can be filled by collecting data from text and building the schema from the bottom up. Meanwhile, the existing schema enables the reader to make predictions and inferences that fill in slots not explicitly outlined in the text. Thus, schema theory provides a theoretical framework for simultaneous processing during reading.

Another excellent paper that dealt with the link between schema theory and reading was that of Anderson and Pearson (1984). The initial portion of their article made three primary points. First, any complete explanation of schemata will have to include information concerning the relationships among the various components. Similarly, such an explanation will have to include a major role for inference. Finally, in order to comprehend language, people must rely on abstract and general schemata as well as their knowledge of particular cases.

Next, the authors approached the role of schemata in inference-making. They described four kinds of inferences which are enabled by schema theory and play a role in reading comprehension. Inferences can be involved in selecting a schema to be used to make sense of a passage. Inference can also be involved in instantiating slots. The third form of inference involves the filling of slots through the assignment of "default values" (Anderson & Pearson, 1984, p. 269). That is, the reader can reason that

since slot X is typically filled by value Y, and there is nothing in this passage to contradict it, Y can be presumed to be true. Finally, inferences allow the drawing of conclusions based upon the absence of knowledge. For example, a reader can reason that "if X is true then it would be stated as such and I would know it to be true, therefore X can not be true." Hence, the presence of appropriate schemata enables the reader to make inferences that move comprehension beyond the literal level.

Schemata also play a crucial role in the allocation of attention according to Anderson and Pearson (1984). It has long been an accepted axiom among researchers that important text elements will probably be remembered more accurately than unimportant elements. Otherwise, comprehension would be completely random. The authors examined three hypotheses that attempted to explain the phenomena. At the conclusion of their analysis they stated: "Despite some inconsistent findings and several unanswered questions, based on the evidence available at this time, the selective attention hypothesis looks promising" (Anderson & Pearson, 1984, p. 278).

As articulated by Anderson and Pearson (1984), the selective attention hypothesis posits four critical points. First, an activated schema and an analysis of task demands provide a test for the importance of upcoming text. Second, each text element is processed to a minimal level and

assessed for importance. Third, those elements which are judged to be important are allotted extra attention. Finally, important text elements are learned better due to that extra attention and are therefore remembered better. Should it prove to be true, the selective attention theory will explain why readers remember critical information better than non-critical information by linking that process to the operations of schemata.

The final section of the Anderson and Pearson (1984) paper that concerns this project involves the role of schemata in remembering. The authors reviewed the research on remembering in light of three hypotheses: a retrieval-plan hypothesis, an output-editing hypothesis and a reconstruction hypothesis. The first explains remembering by positing that schemata provide the structure for a top-down search of memory. The second asserts that schemata provide the basis and motivation for "the selection and rejection of information to report when recalling a passage" (Anderson & Pearson, 1984, p. 281). The final hypothesis states that schemata facilitate reconstructon. That is, a person uses a schema to determine what must have been in the passage based on what actually is recalled and on the expectations generated by the schema itself. Again, the authors reviewed the available literature and concluded "the reader's schema is a structure that facilitates planful retrieval of text information from memory and permits

reconstruction of elements that were not learned or have been forgotten" (Anderson & Pearson, 1984, p. 285).

Taken together, these two articles (Adams & Collins, 1979; Anderson & Pearson, 1984) present a convincing case for the role of schemata in the reading process. Schemata enable simultaneous top-down and bottom-up processing because of their nature as active structures that can be embedded one within another in greater levels of abstraction. These concurrently operating levels further enable slot-filling through inferences made from both available data and the schematic resources of the reader. Schemata also help readers allocate attention to more important bits of information and they provide for purposeful reconstruction during remembering. Hence, schemata provide a general framework for comprehension during reading. One schema that is of great value to readers, and one that has gained much attention, is a schema for stories. It is, therefore, the subject of the next portion of this review.

An important distinction must be drawn between story grammars and story schema. A story grammar is not only an analysis of the structure of stories but also a series of "rewrite rules" which attempt to explain the grammar of a story along linguistically acceptable lines. An example of a rewrite rule would be

Event --> {Episode | Change-of-State | Action | Event + Event} (Rumelhart, 1975, p. 215)

This rule states that an event in a story can be made up of an episode, a change of state, an action, or an event sequence, all of which are mutually exclusive. There has been substantial criticism of these rules (Black & Wilensky, 1979), but the structure of stories which is implied by them has been widely accepted among reading professionals. It is with that structure that this review is concerned.

One of the earliest and most frequently cited treatments of story schema theory is that of Rumelhart (1975). That article was primarily concerned with grammatical rewrite rules, but the basic elements of a story were clearly outlined as well. At the most fundamental level, according to Rumelhart, a story consists of a setting and an episode. The setting is comprised of a series of static propositions that describe the status quo in which the story takes place. An episode is made up of an event and a reaction. Therefore a story must have a setting, an event, and a reaction to that event.

Obviously, however, most stories are more complex than one event and a reaction. Through an analysis of the eleven rules outlined by Rumelhart (1975, p. 219) a specific structure can be described. A story has a setting and is initiated by some event. There is a reaction to that event which includes an internal and an overt response by some character. The overt response includes a plan and some

attempts at executing that plan with some resulting consequence. Subsequent papers on story structure have not substantially disagreed with this scenario (e. g., Mandler & Johnson, 1977; Stein & Glenn, 1979).

Rumelhart (1977) later sought to test his model of comprehension and to show that the gist of a story could be accounted for by that model. He predicted that when asked to summarize a story, subjects' reponses would follow the same general pattern outlined in his model. The position taken by the author is that the "process of comprehension is taken to be identical to the process of selecting and verifying conceptual schemata to account for the situation (or text) to be understood" (Rumelhart, 1977, p. 268).

In support of this stance, the author developed a tree structure that can be used to predict elements of a story that will be recalled by a capable reader when instructed to summarize the story. He applied the structure to five stories and had 10 graduate students read and verbally summarize each one. The summaries were then analyzed with respect to those elements that were predicted by the model. Two statistics are of critical interest. First, of all statements that were made during the summaries, 88% were correctly predicted by the model. Second, of all statements actually predicted by the model, 94% were mentioned by the subjects. To be sure, the experiment employed a small sample that included an atypical population. Nonetheless,

the subjects were good readers and, as such, the evidence they produced is worth noting. That evidence tends to support the claims of schema theorists in general, and Rumelhart (1975, 1977) in particular, that a reader's schema for stories, and the summarization rules that are implied within it, play an important role in comprehension and recall.

In another line of research, investigators have probed two critical questions: Do children have a similar sense of story structure? If so, does it work in the same way? A partial answer was supplied by Mandler and Johnson (1977). This important paper expanded the work on story grammars without drastically altering the structure implied by Rumelhart (1975). The differences in the two structures were limited to two specifics. First, Mandler and Johnson (1977) used the words "goal" and "goal path" which seem to imply more than the broader terms "plan" and "attempts" used by Rumelhart (1975). Second, Mandler and Johnson included in their grammar an ending in addition to Rumelhart's consequence or outcome.

Again, the extensive grammatical and transformational rules developed by Mandler and Johnson (1977) are beyond the immediate interest of this review. What are of importance are the results of their experimentation. To test their theories, they began with 21 subjects from each of the first and fourth grades and from a university. Four stories were

tape-recorded, with each subject hearing and recalling two of them. One recall came after a 10 minute waiting period and the second came 24 hours after hearing the story. Oral retellings were used as the method of recall. Recalls were later transcribed and analyzed.

Results showed that immediate and delayed recall were quite similar, so the data were collapsed for additional analysis. Not surprisingly, adults recalled more than the fourth graders, who in turn recalled more than the first graders. The youngest children recalled settings, beginnings, and outcomes significantly better than attempts, endings and reactions. The fourth-grade students followed the same basic pattern, but attempts were recalled as well as outcomes. Among the adults, attempts were recalled as well as settings and beginnings, but endings and reactions were still less well remembered. The authors also found that less than 2% of sequence errors involved the reversal of basic events in the stories. Then the authors examined the nature of the additions, or intrusions, made by students during recall. It was predicted that additions would be used to fill in missing slots in the stories. Because the stories were particularly well structured, the results may not accurately test the prediction. However, 28% of the additions were used as predicted which the authors considered to be a substantial proportion.

In general, schema theory would predict that once an accurate schema for stories is developed it should be identical for children and adults and should operate in essentially the same way. The schema would be used to guide recall by providing an appropriate sequence and structure. The schema would also enable the inferences that would fill in any missing slots. Mandler and Johnson's (1977) findings provided tentative confirmation of these predictions. Children and adults recalled stories in much the same pattern and that pattern appeared to resemble the structure of the stories. Furthermore, the recalls followed the sequence of the stories almost perfectly. When subjects added inferences, they often did so to fill in missing or misrecalled elements.

Mandler (1978) further tested her theories in an experiment that utilized standard and interleaved versions of canonical two-episode stories. Ninety-six subjects were involved, 24 from each of second, fourth, and sixth grades, and an adult group. In each group, half listened to four stories that had the two episodes intact with one following the other. The rest listened to four stories that had the events of the episodes interleaved, that is, an event from the first episode was followed by one from the second, then one from the first, and so on. The sequence of events was accurate within the interleaving. After 24 hours subjects were asked to retell the stories that they had heard.

Responses were tape-recorded and analyzed in three categories: quantity of recall, quality of recall, and sequence of recall.

For the interleaved stories, adults recalled more than children but the three age groups among the children recalled similar amounts. For the standard stories, there was improvement from second to fourth grade but not thereafter. More entire episodes were left out of the standard stories. In general, all age groups recalled settings, beginnings, attempts, and outcomes better than reactions and endings, replicating the previous findings (Mandler & Johnson, 1977).

The quality of recall was determined by examining additions and distortions. Significantly more distortions occurred when retelling stories that were interleaved during the original presentation. Those distortions included the repetition of nodes, character and event confusions, structural additions, and irrelevant or wrong material. The pattern was consistent across all age groups. Additions were approximately the same for both presentation formats.

During the retellings of interleaved stories, there was a marked attempt, especially among children, to reorder events into a standard, two-episode structure. This seemed to indicate that children's sense of how a story should be structured strongly influenced how it was recalled. In addition, attempts and outcomes were more frequently

recalled together than not, despite their separation during presentation.

The data presented in Mandler (1978) offered strong support for the effects of story schema on recall. Schemata seemed to direct retrieval and determine sequence in order to facilitate understanding and reconstruction of stories. The patterns of recall for children and adults were strikingly similar as were the types of distortions introduced into the retellings of interleaved stories. This study, then, tended to support and expand the positions taken in the earlier paper (Mandler & Johnson, 1977).

Kintsch (1977) also examined the role of schemata in the recall of stories. His discussion of the function of macro-structures in comprehension was quite similar to schema theory and he appeared to use the terms interchangeably. His concern was with determining what in a story signals to the reader that the preceding material should be processed under one category and upcoming material in another. His answer was that the formal and linguistic cues certainly play a role but the more important cues were those that flow from the plot, such as changes in time, location, characters, etc. In other words, the structure of the story itself played the critical role.

The author then referred to four rules based on the work of van Dijk (1977) which attempted to explain how macro-operators help produce summaries of stories. In

effect, the rules stated that irrelevant and redundant material would be deleted, the various members of a given category would be generalized to the category itself, and a sequence of actions would be reduced by introducing a name that refers to the sequence as a whole. Using these rules and the author's theory on the operations of macro-processors, Kintsch (1977) designed a series of studies. In the first, two stories, one well constructed and the other poorly constructed, were analyzed and tested with 25 adult readers. The stories were typed with one paragraph to a page. The subjects were asked to read the paragraphs and to cluster them in any way that seemed appropriate. The subjects' clusterings were quite similar to those which were predicted by the author. He concluded that "the structural information that can be obtained from paragraph sorting experiments is in excellent agreement with the theoretical predictons" (Kintsch, 1977, p. 47).

Next, subjects read each whole selection and rated each sentence in importance. Results showed that significantly more sentences from the well constructed story were rated as important than from the poorly constructed one. Subjects were then asked to write summaries of the two stories. Idea units which had been previously identified as being important appeared more often in the summaries than less important sentences.

Kintsch then reasoned that if schemata played such a critical role in organizing a story, in assessing the importance of data, and in compiling summaries, then when no schema for a particular story existed comprehension should break down. He tested this by selecting four Alaskan folktales for which his subjects should have had no schema due to the culture-specific nature of the stories. He began by asking 40 students to read each sentence in isolation and rate them for comprehensibility. Based on these ratings, they appeared to be completely understandable. Then 36 different students were given four stories to read and summarize. Two of these stories had familiar schemata while the other two were selected from the Alaskan stories. Thirty-six additional students rated the summaries for completeness. Results showed that summaries of stories with the familiar schemata were consistently rated as being better than the others. From this, the author concluded that students' comprehension is enhanced when they have a schema available and it is inhibited when they do not.

To further test the role of schemata in retrieval, Kintsch had students do serial retellings of two stories, one with and one without a familiar schema. In this technique, one student read a story and retold it to a second. The second in turn told it to a third and so on through five retellings. During the final retelling, subjects included significantly more macro-propositions from

the familiar schema story. In fact, the best of the final retellings of the unfamiliar schema stories was considerably poorer than the worst of the familiar ones.

The final study reported in Kintsch (1977) examined the ability of four-year-olds to comprehend identical material both within and outside of a story context. The researcher presented 16 four-year-olds with a three-step task. Using a picture-story, children saw all of the pictures at the same time and their spontaneous comments were recorded. Then the pictures were shown one at a time and children were asked to tell a story with them. Finally, the pictures were taken away and the child retold the story. The experimental manipulation was such that in some instances the pictures were shown in order and in other instances they were scrambled. Results indicated that when material was presented in normal order children did a much better job of telling coherent stories and recalling those stories.

The net effect of all these studies reported by Kintsch (1977) seems to be consistent with the theoretical role of schemata in the comprehension of stories. Kintsch demonstrated that readers organize material along schema-theoretic lines, and that when a schema is available for a story, readers comprehend and summarize it better than when one is not available. Further, when a schema is available much of the original story is retained throughout serial retellings; without a schema the retellings

disintegrate. Finally, students do a better job telling stories from pictures, and recalling them later on, when the order of presentation maintains an acceptable story sequence.

Thorndyke (1977) also tested the effects of story structure on recall. He reported two experiments. In the first, subjects were presented with passages that exemplified one of four possible levels of text structure from a stereotyped level to a random event sequence. The structure was further manipulated by randomly sequencing the sentences within each plot element in half of the presentations. That is, the plot structure was maintained but sentence order within that structure was manipulated. Subjects were given three tasks. After presentation of a passage, subjects were asked to write the passage from memory as close to verbatim as possible. This was repeated for a second passage. Then subjects were asked to write from memory a short summary of each. Finally, they were given a recognition test which examined their ability to determine whether a statement was verbatim from the text or a valid inference based on what was said in the text.

Not surprisingly, the passages with normally ordered sentences were generally more comprehensible and better recalled than the passages with random sentence order within the story structure. However, an important exception was found. When no narrative structure was present, recall of

the normally sequenced sentences and the randomly sequenced ones was not significantly different; without an identifiable story structure being present, sentences may just as well be presented in random order. Furthermore, in the presence of a clearly defined story structure, summaries tended to focus on the central elements of the story and ignore lower level details. Without the structure, summaries were longer and focused on details.

The second experiment reported by Thorndyke (1977) further tested the effects of structure and content variables on the comprehension of adult readers. Two passages were selected for use in the study, one with a very good structure and one with a more ambiguous structure. In addition, one passage had high-imagery characters (a farmer, a dog, etc.) while the other had low-imagery characters (populists and federalists). The two passages were then rewritten into four stories by interchanging the story structures and character sets. Therefore, the four patterns were (a) high-image and complete structure, (b) high-image and incomplete structure, (c) low-image and complete structure, and (d) low-image and incomplete structure. Students were asked to read the stories; to rate the stories for comprehensibility, imagery, and meaningfulness; and to recall them in writing. Results clearly demonstrated that comprehensibility, as judged by the subjects, was a function of story structure while imagery was a function of story

content. Furthermore, recall of the tight structure and high-imagery characters was reliably better than for loose structure and abstract characters. It would appear, then, that this study again demonstrated the powerful effects of story structure on students' sense of the comprehensibility of written material and on their recall of that material.

The last paper to be considered in this review of schema theory and the comprehension of stories is that of Stein and Glenn (1979). The authors had several objections to Rumelhart's (1975) grammar and proposed several alternatives. However, the basic structure of a story according to Stein and Glenn (1979) remained essentially the same as suggested by Rumelhart (1975) and Mandler and Johnson (1977). Their version of a story included a setting and an episode system. That system included an initiating event, an internal response, and a plan sequence which had its own pattern of a goal, an attempt, a resolution, a direct consequence, and a reaction to the consequence. Implied in this, as in all story grammars, is a character that experiences the initiating event, has the goal, and so on.

The authors also devoted much attention to the various links that explain the sequence of events in a story. They noted four varieties: A ALLOWS B to happen, A can CAUSE B,

A THEN B in a specific temporal sequence, and A AND B in a loose linkage in which there is no absolute connection.

To test their theories, the authors reported two experiments in considerable detail. The first had as its goal "to collect recall data in several stories representative of those found in children's literature" (Stein & Glenn, 1979, p. 72). They hypothesized that children should recall information concerning the basic sequence of events, that they should chunk information according to distinctions made in the proposed grammar, that the saliency of categories in recall should vary, and that violations of the temporal sequence would force children to reorganize material during recall. They also expected that recall protocols would contain information defining the logical structure of the sequence of an episode and would contain new information (additions) when certain categories were omitted from the original.

Their sample included 48 children equally divided as to sex and grade level (first and fifth). Four stories were used and children within grades were divided with half hearing stories 1 and 2 while the rest heard stories 3 and 4. Stories were counterbalanced during presentation. The examiner read the story to one child at a time. After an intervening counting exercise, the child would orally recall the story as precisely as possible. The process was then repeated for the second story. All responses were

tape-recorded for future analysis. One week later the child was asked to recall both stories again.

On accuracy of recall, the results were not startling. Fifth graders remembered more than the younger children and more was recalled immediately than after the delay. Data were then grouped by the categories in the grammar. These results were quite variable with only internal responses having a significant grade effect across all stories, with fifth graders recalling more of them. The saliency of individual statements was tested and correlations proved to be quite high both for grade (.84 to .98) and time (.91 to .99). This indicated that the recall of individual items is highly stable. The results concerning the saliency of categories were generally consistent with that of Mandler and Johnson (1977) and Mandler (1978). Major settings (Mandler & Johnson used the term "setting") were best recalled with direct consequences (outcomes), initiating events (beginnings), and attempts close behind. The poorest recall was on internal responses which were subsumed under Mandler and Johnson's reaction category and which they too found to be the least recalled. Relative saliency did not differ with time or age. In effect, then, this portion of Experiment 1 provided replication and substantial confirmation of the results of Mandler and Johnson (1977) and Mandler (1978).

In their analysis of students' additions, the researchers reported a striking finding. The category most frequently added to the recalls was internal responses, the same category which was least accurately reported. This may imply that children are aware of internal responses, but their comprehension of them did not match the researchers' analysis of the original version. Overall, the type of information added was fairly consistent across stories and grades. Fifth graders, however, included three times as many additons of internal responses.

During recall, subjects' sequence of information was very highly correlated with the sequence of the original (.92 to .99). Only one correlation was outside of that range (story 2, first grade, immediate recall) and that corrected to .99 during delayed recall. All but one correlation was higher during delayed recall. This suggested that during recall the structure of the original exerts a powerful influence and that such influence strengthens with time. Furthermore, of those reversals of sequence that did occur, 75% were predicted by the tree structures diagrammed by the authors (Stein & Glenn, 1979, pp. 84-87) and based on their grammar. This suggests that the few reversals which were made were the result of the influences of the child's internal sense of story structure.

The second experiment reported in Stein and Glenn (1979) was intended to examine additional aspects of

processing that could not be assessed with oral recall measures. Two tasks were designed to test (a) children's sense of the types of information that was most critical to a story, and (b) their understanding of causal relationships within and between episodes. This experiment employed 24 children equally divided by grade (1 and 5) and sex. They used the same four stories as in Experiment 1. Children were tested individually and each heard all four stories. After hearing each story the children engaged in the counting exercise and then were asked to tell the examiner the most important thing that had happened in the story. After each child replied, the examiner asked for the second, and then the third most important thing. After all stories were heard, a set of probe questions was asked for each of the four stories to determine the children's sense of the causal relationships.

The results of Experiment 2 confirmed the importance of internal responses and the developmental nature of their inclusion. In assessing relative importance of items within a story, first graders included consequences more than any other category, with internal responses second and attempts third. Fifth graders cited internal responses most often, with initiating events and consequences second and third. It should also be noted that 25% of all importance judgments were inferred.

During the probe questions children had little difficulty answering correctly. They seemed to be fully aware of the causal relationships both within and between episodes. Furthermore, internal responses were clearly perceived as being the primary causal factor.

In summing up their research, the authors stated that they had "presented and partially validated a schema for stories" (Stein & Glenn, 1979, p. 115). It is difficult to take exception to that conclusion. Their evidence uniformly supported the vital role that schemata play in the organization of story information in memory and its recall.

The research cited in this review of schema theory has suggested that story schema functions in essentially the same manner for children and adults. However, the evidence also revealed several important differences in the quantity and quality of recall among the various age levels that were studied. Mandler and Johnson (1977) demonstrated that the quantity of recall increases from first to fourth grade and from fourth grade to adult. In addition, they demonstrated that each of the age levels has a slightly different pattern of category saliency during recall. Mandler (1978) confirmed these differences and found that children of various ages respond to interleaved stories with differing levels of attempts to reorder events during retellings.

Stein and Glenn (1979) also found age-related differences in children's ability to use their story schema.

Fifth graders, for example, included significantly more internal responses in their retellings than first graders, and the two groups differed in the kinds of information they cited most frequently as being important to a story.

These differences in children's use of story schema suggest that such usage is developmental in nature. This conclusion has been supported by Hoover (1982). She studied children's ability to detect incoherence in narratives. Her subjects included 112 children enrolled in a Head Start program, in kindergarten, and in first grade. She read the same story to the children six times using three normal versions of the story and three versions with some structural alteration. The alterations included the reordering of the setting, the initial event, and the consequence to various parts of the story. She then tested the children in the fall and the spring of the school term for their ability to detect which passages had the incoherent structure. The results showed that significant growth occurred for both the fall and spring administrations between kindergarten and first grade, and that significant growth occurred from the fall to the spring within each of those grades. These developmental differences led the author to conclude that "even at the earliest stages of reading acquisition, story schema exerts a significant influence on the processor's ability to monitor narrative material for incoherence" (Hoover, 1982, p. 272).

The developmental nature of story schema suggests that instruction in the basics of story structure may have a significant effect on children's comprehension. Several authors have researched this possibility with generally positive results.

An early article concerning the use of story schema as an instructional device came from Cunningham and Foster (1978). They reported that Foster developed a simplified version of a story structure which included setting (location, time, and characters), theme (main goal), plot (subgoals, attempts, and outcomes), and resolution. She used it in her sixth- and seventh-grade classes and found that her students were better able to understand the story which they were studying.

Unfortunately, Cunningham and Foster (1978) provided only a testimonial as to the instructional potential of story structure. They produced no empirical evidence to substantiate their claim. Dreher and Singer (1980) tested the strategy and found it ineffective. They established three groups: a story grammar group, a control group that read all of the material, and a social studies group that received content area instruction. The students in the experimental group received a three-step strategy using a model story, small group activity aimed at analyzing a second story, and individual analysis of a third. After the treatment period, the duration of which was not defined by

the authors, all students performed a written recall of a fourth story.

Results from the experiments were mixed. Experimental group students learned to categorize information according to the schema and when other students were asked to try the same task they did significantly worse on it. When comparing the number of propositions recalled during the written posttest, however, there were no differences among the three groups. All students recalled propositions from high in the story grammar hierarchy. Based on these results the authors concluded that students can learn to categorize information from a story grammar perspective; that ability, however, is not translated into better recall.

There are a number of factors that could have accounted for the results obtained by Dreher and Singer (1980). The posttest material and its manner of presentation (it was read to students while they followed along) might have rendered the test so easy that it lost its capacity to differentiate between levels of recall. It is also possible that the instructional treatment did not affect recall at all because of the apparent shortness of the duration of the experimental period. Perhaps enough time was not allowed for transference from categorization with text to recall without it, two very different skills. A third possibility is that the schema for stories of the fifth graders used as subjects in the study was already well developed. If that

was true, the posttest would have resulted in few observable differences. In any event, their results seemed to contradict the testimonials of Cunningham and Foster (1978).

A later study in which Singer was involved, however, appeared to reach an opposite conclusion (Singer & Donlan, 1982). They taught eleventh graders to derive schema-related questions from the stories they were reading and to use them as guides to comprehension. The subjects read six stories over a three week period. Comparisons with a control group showed significant differences in comprehension in favor of the treatment group. The authors concluded that instruction such as they offered, which utilized a story schema approach, helped to increase comprehension. They also concluded that the simple story grammars children develop in their early school years may be effective for understanding simple fables, but more complex schemata are needed for reading complex short stories.

Another attempt at investigating the effects of direct instruction in story structure was reported by Fitzgerald and Spiegel (1983). They worked with 19 fourth graders who were identified as being particularly weak in their knowledge of story structure and who performed poorly on a standardized test of reading skills. Subjects were randomly assigned to treatment and control groups. The control group was instructed in dictionary skills. The experimental group received a two phase instructional program. First, students

received six 30- to 45-minute sessions spanning a two-week period. The second phase included 10 sessions of the same duration over a five week period. During the initial stage, each lesson focused on one specific element of story structure and its typical sequential position. In the latter stage, individual and group lessons were designed to reinforce knowledge of story structure and to increase awareness of the relationship between story structure and comprehension.

Six different analyses were performed with posttest data. A story production task, in which students wrote a story after being given a setting, was used to assess students' knowledge of story structure and temporal order. A scrambled stories task, in which students read a scrambled story and retold it a day later, assessed knowledge of story structure through students' spontaneous reordering of events during retelling. Protocols were also scored for total recall, additions, and distortions. A test with 17 literal and inferential comprehension questions measured recall and understanding.

The results of the analyses showed several positive effects for the treatment group. Knowledge of story structure was greater among experimental group students as was their ability to make inferences and to integrate information. Experimental students also had a greater awareness of the appropriate sequence of story elements

within the spontaneously generated stories and had greater accuracy of recall for the scrambled stories, even after a 24-hour delay. Overall, the authors concluded that instruction in story structure was useful in enhancing reading comprehension.

Fitzgerald and Speigel's (1983) evidence is in direct contrast to that of Dreher and Singer (1980) and tends to support the claims of Cunningham and Foster (1978). It seems to indicate that instruction in the structure of stories can significantly improve the ability of below average fourth-grade readers to understand that structure and to comprehend stories.

An alternative to instruction in story structure itself is a questioning strategy based upon that schema. Bowman (1980) and Bowman and Gambrell (1981) reported the results of a study involving 100 sixth graders. Students were randomly assigned to experimental and control groups within three different levels (good, average, and poor readers). Passages used in the study were rewritten to approximate the independent reading levels of each group. The treatment groups received a post-reading strategy involving questions constructed according to the categories of a typical story. The control group students were asked questions that were more traditional and resembled those found in basal readers. Two types of posttests were used. One involved free recalls and the other used cued recalls. Results indicated that

experimental group students performed significantly better on both posttest measures. There was no treatment by level interaction so it would appear to be safe to conclude that the strategy worked equally well at all ability levels.

In summation, a number of theorists have posited the existence of psycholgical constructs called schemata which help people make sense of new experiences. Schemata provide an organizational framework for classifying new information by comparing it to previous experiences. When an experience is entirely unique, the schema-making process builds the ideational scaffolding needed to interpret and understand the new concept. Schemata are believed to be active, evolving constructs that operate both from a top-down and a bottom-up perspective and can be embedded within each other to greater levels of abstraction.

Theorists in the area of reading research have suggested that one of the many schemata that all persons possess is a schema for stories. This schema provides a set of expectations for the reader based on the perception that a story has certain elements (a setting, a character, an initiating event, etc.) which can reliably be supposed to occur in any given story. Using this information, readers search for these elements, and in their absence make a series of inferences that fill in the missing slots. Hence, comprehension is enhanced beyond the strictly literal level. Research in the latter half of the 1970s and the early 1980s

has generally tended to confirm predictions made on the basis of story schema theory.

Research on instructional uses of story structure has not been as consistent. Some authors have found positive effects for instruction in story structure while others have not, primarily Dreher and Singer (1978). However, the preponderance of evidence suggests that a schema for stories provides an efficient method for classifying what students recall after reading or listening to a story. Furthermore, it seems likely that readers use their sense of a story's structure to facilitate comprehension during reading and during retrieval and reproduction of information afterwards.

For the purposes of reading practitioners, and based on the preceding review, the primary utility of story schema theory seems to lie in its potential use as a framework for diagnosis. It appears to be possible that the analysis of students' recall of stories along story structure lines could be an efficient diagnostic technique. Many of the studies reviewed above successfully used oral recall protocols, or retellings, to tap students' comprehension of stories. Research examining the use of retellings in analyzing and enhancing students' reading comprehension is reviewed next.

Part II: Using Retellings in the Classroom

The use of free recalls, both oral and written, as a method of assessing reading comprehension gained some initial attention during the second decade of this century. It won few proponents, however, and the method generally diasppeared until its re-emergence in the 1970s as a research tool. Then, in the early 1980s, it attracted some attention as an informal assessment procedure for classroom use. This review will summarize the initial work completed by Courtis (1914), H. A. Brown (1914), and Starch (1915) before focusing on the recent interest in retellings for classroom use.

Courtis was known primarily as a developer of tests in reading and language arts (e.g., Courtis, 1914, 1917). He contended (Courtis, 1914) that reading comprehension could be measured effectively by a technique which involved the use of written retellings. The examiner would select an appropriate passage and count the number of idea units contained within it. The subjects would read the passage, set it aside, and write out everything that they could remember, trying to reproduce the original as accurately as possible. Next, the examiner would count the number of idea units in the retelling that accurately reproduced the original. The resulting proportion would yield a measure of the subjects' comprehension of the passage.

H. A. Brown (1914) also advocated an approach to comprehension assessment that involved written retellings. Brown believed that three variables should be considered when measuring comprehension: (a) rate of reading, (b) quantity of reproduction, and (c) quality of reproduction. The first of these was measured by giving subjects a lengthy passage to read and determining the number of words read in one minute. The second measure, the quantity of reproduction, involved determining the proportion of idea units remembered in the same manner as suggested by Courtis (1914).

The final measure was somewhat more complex. The examiner had to compare each idea unit in the reproduction to the original. Then, only those units were counted that accurately and completely reproduced the original in every respect. The author noted that the exact words of the original were not necessary for the reproduction to be counted as wholly correct. This final proportion constituted the quality of reproduction measure according to H. A. Brown (1914).

A third approach to assessment using free recalls was advocated by Starch (1915). In his method, the unit of interest was the number of words in the reproduction which accurately reflected the original. The examiner was expected to read the reproduction and cross out any words that were incorrect, that added new ideas, or that repeated

other ideas already reproduced. In the author's opinion, the remaining number of words provided an index of comprehension.

These three techniques are somewhat inefficient in that they suffer from a high degree of ambiguity. Furthermore, the authors reported no empirical evidence in support of their assertions.

Courtis (1914) did not define what was meant by an idea unit nor what constituted an accurate reproduction of an idea from the original. Neither did he provide a scale by which the acceptability of the subject's work could be measured. As a result, the procedure is vague and would seem to be of minimal use to teachers. Nonetheless, the approach appeared to have some value and is similar in many ways to the research tools of Kintsch and van Dijk (1978) and Turner and Greene (1977).

H. A. Brown's (1914) method added the element of reading rate and differentiated between quantity and quality of reproduction. However, several factors were left unexplained. The quantity measure retained all of the ambiguity of Courtis' (1914) method. The quality measure was more precise in its definition of an entirely correct reproduction but still did not define what constituted an idea unit in the original passage. Finally, Brown did not describe the interaction among the three measures. As a

result, the scores would appear to have little value to a classroom teacher.

The third approach presented by Starch (1915) raises a number of questions. Would an accurate inference which added some details not in the original be crossed out as "adding new ideas"? If not, wouldn't it be possible for the reproduction to have more acceptable words than were contained in the original? This could result in a proportion of recall in excess of 1.0 which, of course, is nonsensical. If such an inference should be stricken from the reproduction, the subject would be rewarded only for literal comprehension and perhaps penalized for including an accurate inference.

Questions such as these and that of Kelly (1916), who argued that comprehension and reproduction are very different skills and that the former should be measured with minimal interference from the latter, resulted in the almost complete disappearance from the research scene of the use of retellings in assessment. Johnston (1984) gives credit to Y. Goodman and Burke (1970) for reviving the use of free recall protocols and referring to them as retellings. However, he concluded that the procedure requires too much time for widespread use.

While this is certainly true of the labor-intensive analysis used in research by Kintsch and van Dijk (1978), it is not necessarily true of the more recent techniques

developed for use by practitioners. The next portion of this review, therefore, focuses on the use of retellings in the classroom.

An early examination of the use of retellings in the classroom was reported by Zimiles and Kuhns (1976). One portion of an extensive study of children's retention of narrative material involved the use of immediate retellings after listening to a story. A total of 526 kindergarten children participated, half of whom did immediate retellings. On recall tasks delayed two days, two weeks, and two months, children who had performed the immediate retelling scored consistently higher. Based on their extensive data, the authors concluded: "The opportunity for children to recall the story immediately upon hearing it very substantially facilitated subsequent memory of it. The retention of children with immediate recall exceeded that of children without immediate recall under virtually all experimental conditions" (Zimiles & Kuhns, 1976, p. 17). In fact, the authors noted that seven-year-olds with immediate recall performed better under nearly all experimental conditions than eight-year-olds without it.

This study presented some solid evidence for the positive impact of retellings on recall of narrative material. Unfortunately, most of the articles on the topic that appeared during the eight years following its publication added little to the research base concerning

classroom uses of retellings. Instead, most pieces were non-research oriented "thought pieces" which proposed various procedures for using the technique in the classroom without producing evidence in support of the claims.

Clark (1982, 1984), for example, reported a technique that he called the "Frecall Method." It involved breaking a passage into "pausal units" by placing a slash wherever a good reader would tend to pause when reading aloud. Next, the teacher would assign a number from one to three to each unit according to its perceived importance to the selection. Then the student would read the passage and retell it, including everything that could be remembered, while the teacher tape-recorded the retelling. In order to analyze the protocols, the teacher should first determine the sequence in which the pausal units were recalled and whether or not that sequence was acceptable. Then the proportion of units recalled would be calculated as would their mean importance level.

The information gained with this technique can be valuable to a teacher. After some experience with it, the teacher would begin to sense what proportion of units recalled and what mean importance level would indicate an acceptable level of comprehension. Also, the appropriateness of the sequence of the student's recall would be evident. The author added a caution that these bits of information should be considered together rather

than in isolation. It would be possible, for example, for a student to recall a very high proportion of low level details. As a result, the overall proportion would be high but the pattern of comprehension unacceptable. Taken together this information may be quite enlightening to a teacher.

There are at least two objections to the Frecall Method that could be raised. First, the actual statistics produced by the method are vague and reveal very little usable information to the teacher. Knowing that a student recalls 32% of all pausal units does not help a teacher plan a lesson or isolate specific areas of weakness. Similarly, a mean level of importance of 1.9 may mean that the student and teacher disagreed as to what was important to the story but would not, in itself, be of great value. In the example given by Clark (1982, p. 437) the subject's mean importance level was reported as 1.9 which Clark determined to be weak. Yet, based on Clark's own assessment of the importance of each unit, if the subject remembered and accurately reproduced every single pausal unit in the selection the resultant mean importance level would be slightly over 2.0. Surely perfect recall could not be considered weak. Nonetheless, the example given by Clark would tend to indicate that it would. In fact, the directions to the students to include everything that could be remembered may tend to inflate the mean importance level by encouraging

students to include minor details that would not ordinarily be included if the directions were to summarize the story.

The second objection to the Frecall Method concerns the preparation required to use the technique. While it certainly is not as labor intensive as the use of free recall protocols in research, it still requires that the teacher analyze a story one unit at a time and type it out by units on a form. During analysis the teacher would have to examine each statement of the subject and check off a corresponding unit in the original. Some students may make statements that imply knowledge of other units not explicitly mentioned. This too would have to be analyzed and the appropriate units checked on the form. Furthermore, this level of analysis may preclude the use of full-length basal stories. Thus, the teacher would have to search for appropriate short selections, which is another time-consuming chore. With all of these time constraints in mind it is possible that teachers would avoid using the technique on a continuing basis.

Overall, the Frecall Method would seem to have great potential as a diagnostic technique, but its inability to deliver classroom oriented information and the time required to apply it may limit it to only occasional use. It would not appear to be the type of technique that can become a normal part of reading group instruction.

A major criticism of assessing oral retellings was voiced by Irwin and Mitchell (1983, 1984). They stressed that systems such as those advocated by Y. Goodman and Burke (1970) and Clark (1982) do not deal with the interrelationships among the various elements of a story and the quality of their reproduction by a subject. What was needed, according to the authors, was a method that could evaluate retellings in an integrated, holistic fashion. Accordingly, their technique was aimed at judging the richness of retellings.

The authors began by establishing five levels of richness for a retelling. Level 5, the richest, was described as follows: "Student generalizes beyond text; includes thesis (summarizing statement), all major points, and appropriate supporting details; includes relevant supplementations; shows high degree of coherence completeness, comprehensibility" (Irwin & Mitchell, 1983, p. 394). Level 1, the poorest, was described as follows: "Student relates details only; irrelevant supplementations or none; low degree of coherence, completeness, and comprehensibility" (Irwin & Mitchell, 1983, p. 394).

To use the technique teachers would ask students to read a passage orally and retell it. The teacher analyzes the retelling according to eight categories. The presence, absence, relevance, or degree of excellence of those categories would be assessed and checked on the matrix

provided by the authors. Ultimately, the pattern of the assessment would result in an overall rating of the richness of the retelling.

To test the inter-rater reliability of their process, 10 teachers assessed the transcriptions of 48 retellings done by 24 students, each one reading a narrative and an expository text. After initial instruction in the technique teachers' agreement on the rating scale reached only 38%. A second instructional session followed, this time with specific examples of retellings at each level of richness. Teachers were then asked to re-examine the same 48 retellings. Agreement after the second training session reached 87.5% with total agreement being achieved when ratings one level below and above the authors' own ratings were included.

As a result of their preliminary investigation, Irwin and Mitchell (1982) reported three observations: (a) that retellings can be scored holistically, (b) that their technique can be used with either narrative or expository texts, and (c) that teachers can use the technique after adequate training.

There are several points of interest concerning this technique that should be mentioned. First, since little effort is required of the teacher in advance, this method would appear to be cost effective in terms of time and effort. In fact, Leys (1984) reported that scoring a

retelling according to the holistic method of Irwin and Mitchell (1982) required only about 40 seconds per retelling. It would seem that such a small investment in time would prove to be worthwhile in terms of the information gained and the opportunity for eliciting student talk.

Second, the richness scale directs the teachers attention to more than the reproduction of certain phrases from the original. The inclusion of skills such as the ability to genralize, summarize, or supplement text can help teachers think about comprehension beyond the level of literal recall. These two factors tend to indicate that the richness scale is a technique worthy of further investigation.

Like the Frecall Method (Clark, 1982), however, the richness scale raises some as yet unanswered questions. First, it is not clear what a rating of 3 would mean to a classroom teacher concerned with teaching comprehension. No specific comprehension errors or weaknesses are pinpointed so little direction for immediate instruction is provided. Second, the inter-rater reliability study cannot be accepted without question. It is not clear, for example, what impact the teachers' knowledge of the initial low reliability and their rescoring of the same protocols could have had on the eventual outcome. It is at least possible that subjects who rated the retellings consistently high the first time

lowered their ratings while others who consistently rated them low increased the scores. This would yield an artificially high level of agreement. It would have been more accurate to train several new raters using the second training method and determine the new level of reliability without the confounding factors being present.

A partial answer to the question concerning the usefulness of the technique was provided recently by the same authors (Irwin & Mitchell, 1984). They developed a second checklist to be used in conjunction with the first. This second checklist included greater specificity about readers' comprehension of text, metacognitive awareness, and facility with language. Unfortunately, no information on the training or time required to use the checklist, its reliability, or its use by teachers was provided. Therefore, research on its effectiveness and efficiency will have to be performed before any judgments can be made. In general, however, it appears to be a serious departure for the authors from their richness approach in favor of greater specificity and utility.

Another effort at specificity and utility with respect to retellings has been reported by Marshall (1984b). Initially (Marshall, 1980), she developed a series of questions based on story structure theory (Mandler & Johnson, 1977; Stein & Glenn, 1979). The questions were designed to probe students' recall of stories by focusing on

story elements such as setting, theme, and characters. Later, Marshall (1983, 1984b) added a retelling checklist using the same elements.

On the checklist, the story elements were listed across the top and students' names were listed along the left side. Students were asked to read a passage and to retell it. If the student mentioned a particular story element, the teacher placed a plus sign in the appropriate box. Next, the teacher used the probe questions to examine any area not mentioned during the retelling. If the student responded correctly, a check mark was placed in the box. A minus sign was used to indicate that even after the probes a category remained unmentioned or incorrectly answered. With this method, a pattern of recall, both spontaneous and cued, emerged which could guide further instruction.

In the opinion of the present researcher, Marshall's (1983) approach to assessing retellings has much to recommend it to teachers. It can be used with basal stories of any length and it takes advantage of children's sense of story structure. As such, it offers instant feedback relative to the student's immediate task at hand. It allows teachers to evaluate their students' success with the materials they must use on a continuing basis rather than introducing some extraneous material that may not resemble the reading tasks of the classroom. Furthermore, the probes

allow teachers to examine students' recall of areas of the story not mentioned during retelling.

These advantages can make Marshall's technique a valuable asset to teachers. However, there are several factors that could limit its use. First, the author, like Clark (1982) and Irwin and Mitchell (1983, 1984), did not present any empirical evidence in support of her claims. Second, the plus, check, and minus system does not work well with multiple episode stories. The teacher is instructed to use a plus if the student mentions a given category during retelling. How should that be interpreted if a student retells some significant events, mentions others in response to probes, and leaves others out completely? Should the teacher place a plus, check, and minus in the box? It would appear that at least for longer stories, some gradation would be helpful within each of the categories.

The third possible limitation is probably less significant than the others, but it could restrict its use by teachers. Marshall (1983, p. 619) wrote that teachers occasionally may not want to ask questions about a story or have the time to do so. She offered her checklist as a solution to the problem. Nonetheless, to use the checklist according to the author's directions the teacher must use the probe questions. This does not solve the problem that Marshall tried to address. It is at least possible that the extra demands on teachers' time and the necessity of

spending that time with one student may inhibit teachers from using the technique on a continuing basis. The extra time may be time well spent (the data have yet to be reported) but it may not be available.

In summation, Marshall's (1983) method of using a story structure approach to assessing a retelling holds great promise. It offers teachers a relatively easy method of assessing a student's comprehension and can be used during a typical reading lesson with basal materials. The questions raised above may suggest some modifications to the procedure, but they do not dispute its basic premise.

Recently, Leslie Morrow reported some empirical evidence concerning the effects of retelling stories with kindergarten children. In a series of articles (Morrow, 1984, 1985a, 1985b) and a presentation (Morrow, 1985c) she reported the results of two studies that have implications for this investigation. In the first study (Morrow, 1985a, 1985b, 1985c), she developed a comprehension pretest and posttest comprised of five traditional questions and five story structure questions. Then she randomly assigned 29 children to an experimental group and 30 to a control group. A story was read to all students and the treatment group students immediately retold the story to an adult. Control students illustrated a portion of the story that they thought was important. An analysis of covariance revealed

that the treatment group performed significantly better on the posttest than the control group.

In the second study, Morrow (1984, 1985a, 1985b, 1985c) collected data on 76 students, 32 of whom were randomly assigned to an experimental group and 44 to a control group. Once each week for eight weeks a story was read to all subjects. As in the first study, the experimental group immediately retold the story while the control group illustrated it. In this study, however, teachers had a retelling guidesheet and assistance was offered as needed during the retelling. Prompts were in the form of questions or instructions about structural elements of the story. A sample prompt would be "Tell how the main character solved the problem." All children took a comprehension pretest and posttest similar to that used in the first study, and performed a pretest and posttest retelling. Retelling tests were given without the use of the guidesheet.

Results showed that experimental group students made significant gains over the control group on traditional questions, story structure questions, and on total comprehension scores. Experimental group students also performed significantly better on the retelling posttest, especially with respect to theme, resolution, sequence, and total score.

In these two studies, Morrow replicated a portion of the Zimiles and Kuhns (1976) research by demonstrating that

when kindergarten children retell a story immediately after hearing it, their comprehension is enhanced. In addition, she offered preliminary evidence that guided retellings which employ a story structure approach can help young students develop a better sense of story structure. Taken together, the Zimiles and Kuhns (1976) and Morrow (1984, 1985a, 1985b) studies form the beginning of a research base on the use of retellings in the classroom. It would appear that the retelling technique can be more than a research tool; it may also help improve comprehension and may function as an effective assessment procedure.

The assessment capabilities of retellings were further examined by Smith and Jackson (1985). Their approach, however, is markedly different from those already discussed. A written retelling of a specific passage is used as a placement technique by the Learning Skills Center at Indiana University. An 800-word passage was extensively analyzed to to determine and weigh its major generalizations, supporting details, and text structure. Students were instructed to read and to study the passage and were encouraged to make notes and to rehearse information. Then they were asked to reconstruct the passage from memory without reference to the passage or their notes. An elaborate scoring procedure was used to rate the written retelling. The researchers reported a high degree of reliability for the instrument and claimed that their results support the conclusion that

"performance on the test does relate to abilities needed for academic success" (Smith & Jackson, 1985, p. 627).

While this particular use of retellings has no direct application to the classroom, it is cited as a strong example of the effectiveness of one assessment procedure that relies on retellings. In addition, the authors' conclusion that the skills revealed in the retelling relate to the abilities needed for academic success has implications for this investigation. At the very least, it is an encouraging sign to other researchers interested in the uses of retellings.

Until now, most of the data which have specifically addressed the use of retellings in the classroom have been reported by Zimiles and Kuhns (1976) and Morrow (1984, 1985a, 1985b, 1985c) and have involved kindergarten-aged children. Very interesting techniques have been reported by Clark (1982, 1984), Irwin and Mitchell (1983, 1984) and Marshall (1983, 1984b). All of this latter group have reported success with their methods and produced testimonials from practitioners in support of that success. Morrow's (1984, 1985a, 1985b, 1985c) research suggests that the story structure approach when coupled with retellings can be a valuable aid to teachers, and Marshall (1983, 1984b) agrees. Additional evidence in support of that

stance, especially as it applies to the middle grades, needs to be gathered.

Part III: Critical Variables in the Assessment of Retellings

The items to be reviewed in the following pages were all suggested by Judy Mitchell and her colleagues (Mitchell, Clark, Grant, Irwin, Marshall, Mason & Leys, 1984). Throughout their presentation, emphasis was placed on six of the more critical variables which are involved in either children's comprehension of stories or the assessment of comprehension. Those variables were (a) the maintenance of an appropriate sequence during retelling, (b) the analysis of inferences revealed in the retelling, (c) the analysis of the intrusions or additions included in the retelling, (d) the importance of children's monitoring of their own comprehension, (e) the importance of an adequate vocabulary to comprehension, and (f) the importance of an adequate general background knowledge to comprehension.

Two important points about this portion of the review need to be made. First, the items are treated separately for convenience and clarity. In reality, they are so interconnected that it is almost impossible to understand or analyze any one of them in isolation, even if it were desirable to do so. Second, the following reviews are only cursory treatments of very complex issues. The intent is not to review the literature on, for example, metacognition to any great depth. Rather, it is to demonstrate that

metacognition is believed to play a crucial role in the comprehension process and therefore should be examined in a reading diagnostic process.

Sequence

The recall of events and story elements in an appropriate sequence is an important principle both in schema theory and research concerning the structure of stories (Mandler & Johnson, 1977; Stein & Glenn, 1979). For example, based on their theory of story structure, Kintsch and van Dijk (1975) presented readers with alternate forms of two stories. One story followed a typical story structure format while the other included paragraphs whose order within the story was scrambled. Given sufficient time to read and assimilate the story, subjects wrote summaries of the scrambled stories that were indistinguishable from summaries written from intact stories. This suggests that in order to make sense of the scrambled story, subjects re-ordered events into an appropriate sequence.

Further evidence of this re-ordering tendency was provided by Stein and Glenn (1978) as cited by Stein (1979). The authors constructed nine different versions of an expected story sequence by altering the locations of three categories. Subjects were asked to read and recall the scrambled material and in nearly every case subjects spontaneously reordered the categories into a more appropriate sequential pattern. This confirmed the tendency

noted by Kintsch and van Dijk (1975) for readers to construct an appropriate sequence when one is absent, and suggests the importance of a proper sequence to comprehension.

Rumelhart (1975) has postulated that the sequence of events in a story can be connected by four kinds of relationships. Event 1 can ALLOW event 2 to happen; 1 can CAUSE 2; a temporal sequence can be established by a 1 THEN 2 relationship; and the events can be loosely connected in no particular order by a 1 AND 2 relationship. Stein and Glenn (1979) also addressed the sequencing of events and seemed, to the present reviewer, to include the ALLOW function in the THEN category. It has been hypothesized that these varying kinds of relationships should be recalled with differing degrees of saliency (Glenn 1977). She suggested that the stronger link between events in the CAUSE relationship would lead to its being remembered more reliably than events in a THEN chain. She tested her expectations by manipulating the relationships among episodes in a story. Subjects responded by recalling events with the expected degree of saliency depending upon whether they were connected by a CAUSE or THEN relationship.

Based on these results, it would seem to be clear that the sequence of events and categories in a story, and the kinds of relationships that link those categories, are important both to comprehension and recall. Furthermore, in

the absence of an acceptable sequence, readers tend to re-order events for greater comprehensibility. Therefore, the sequence of a child's retelling can indicate to a teacher how carefully that student is attending to the story sequence and the linkages between events. If events are retold in an inappropriate pattern, it is a signal that there is a serious breakdown in the comprehension or the recall process.

Inferences

As is indicated in Part I of this review, schema theorists believe that the ability to make inferences plays an important role in several aspects of reading comprehension. Anderson and Pearson (1984) identified four kinds of inferences that affect reading comprehension including those that are used to select schemata, instantiate slots, assign default values to slots, and draw conclusions based on the absence of certain information.

The importance of inference was demonstrated in a widely known study by Anderson, Reynolds, Schallert, and Goetz (1977). They presented college students with two texts that were open to differing interpretations. One interpretation was closely aligned with students' major area of study (physical education majors and a wrestling interpretation, music majors and rehearsing a quartet) while the other was aligned with a more common schema (breaking out of prison and playing cards, respectively). Physical

education majors tended to choose the wrestling schema and the card-playing schema for the selections. Music majors chose the prison break and the quartet rehearsal schemata. These results strongly suggest that the interpretation of the passages depended to a great extent on inferences arising from students' personal experiences.

Goetz (1979) further demonstrated the role played by inferences in reading comprehension. He was interested in whether or not an inference is more likely to be made based upon its importance to an understanding of the selection at hand. He designed two versions of a story in which crucial information and trivial information were sometimes explicitly stated and implied at other times. He then tested subjects' recall and recognition of both kinds of information. His findings indicated that implied important information was consistently recognized and recalled. When that information was stated explicity, importance predicted recall but not recognition. Goetz's (1979) study provided some evidence that when an inference is critical to a coherent understanding of a story, it can reliably be expected to be made by efficient readers.

An interesting theory of the role of inference-making in reading was presented by Collins, Brown, and Larkin (1980). They noted that in cognitive psychology inference is typically thought of as a slot-filling activity similar to that proposed by Anderson and Pearson (1984). According

to that theory, inferences are made in order to fill in missing connections in the text which is being read. The alternative hypothesis offered by Collins et al. (1980) proposes that comprehension of text is accomplished through building a cognitive model of that text through inferences. The target structure, or model, provides the organizational principles which guide the inferencing procedures. The full comprehension of text proceeds from the early model to progressively more complex models and it is the inferencing process which plays the key role in model refinement.

Whether the primary role of inferences is to fill in missing slots and connections or to guide the process of model refinement, it is clear that inference-making is critical to reading comprehension. As such, it is equally important that teachers assess their students' ability to make needed inferences. Therefore, the checklist used in this investigation directs teachers' attention to any inferences that are included in a retelling.

Intrusions

Intrusions are included in this study more from a procedural than from a theoretical perspective. Intrusions are comments, opinions, or bits of information introduced by the student into a retelling. Often, an intrusion can reveal crucial information as to the quality of a student's comprehension, inferences, vocabulary, or background knowledge.

For example, during a pilot project prior to the present investigation, the researcher asked a fifth grade student to read a story about a girl who rode off to warn the citizenry of an impending attack by the British. The student's retelling included the entire procedure for saddling a horse despite the fact that the procedure was not even alluded to in the story. Another student said the girl rode off "sort of like Paul Revere." Both of these intrusions reveal something about the students who made them. The first student obviously had some prior experiences with saddling a horse and so his comprehension of that portion of the story was enriched. The second student made an accurate comparison between the heroine and Paul Revere which revealed something about the student's knowledge of history and understanding of the heroine's purpose.

A number of studies have successfully used the analysis of additions to reveal which categories of a story's structure are most frequently supplied or inferred by readers (Mandler & Johnson, 1977; Stein & Glenn, 1979) and to provide a partial assessment of the quality of recall (Mandler, 1978). Therefore, it seems apparent that the analysis of intrusions made during retellings can be used as a vehicle for the assessment of students' understanding of text.

Metacognition

According to Baker and Brown (1984), two different phenomena are included under the term "metacognition." Those phenomena are one's knowledge of one's own cognitive processes and one's ability to regulate those processes. The former refers to individuals' knowledge of their capacity to deal with the tasks at hand. The latter refers to the ability to use self-regulating devices during reading, such as checking the correctness of hypotheses or altering reading speed based on the difficulty of the material. Comprehension monitoring, a function of the second phenomenon, is thought to be a necessary part of any attempt to read for meaning (Baker, 1979). In fact, Brown listed seven comprehension monitoring activities that the effective reader uses to ensure successful comprehension:

1. clarifying the purposes of reading, that is, understanding the task demands, both explicit and implicit;
2. identifying the aspects of a message that are important;
3. allocating attention so that concentration can be focused on the major content area rather than trivia;
4. monitoring ongoing activities to determine whether comprehension is occurring;
5. engaging in review and self-interrogation to determine whether goals are being achieved;
6. taking corrective action when failures in comprehension are detected; and
7. recovering from disruptions and distractions. (A. L. Brown, 1980, p.456)

The capacity to use these and other comprehension monitoring skills is apparently age- and ability-related. A study of second-, fourth-, sixth-, and eighth-grade students' perceptions of reading was conducted by Canney and Winograd (1979). They found that children in grades 2 and 4, and poor readers in grade 6, viewed reading as primarily a word identification process. The better sixth-grade readers and nearly all eighth graders believed that the purpose of reading was to obtain meaning from text. Furthermore, poor readers also tended to believe that they could read paragraphs made up of random words because they could pronounce all of them correctly.

Thus it would appear that while comprehension monitoring is necessary when reading for meaning, young and poor readers do not tend to employ self-regulating activities. However, there is some evidence to suggest that with appropriate instruction poor students can improve their monitoring efficiency. Bransford, Stein, Shelton, and Owings (1981) worked with children in both the bottom and top quartiles of the fifth grade. When presented with stories that lacked congruence between a character's description and actions, better readers were able to identify the inconsistencies but poor readers were not able to do so. Subsequent instruction, however, led poor readers to increase their ability to identify incongruencies to a substantial degree.

In summation, comprehension monitoring is believed to play an important role in the process of comprehending during reading. Typically, young readers and poor readers do not spontaneously use monitoring strategies and they view reading as a word-calling exercise. There is tentative evidence, however, that once these failures are identified directed instruction can be used to help alleviate the problem. Therefore, it seems appropriate that a retelling checklist direct teachers' awareness toward evidence of the presence or absence of comprehension monitoring skills in their students.

Vocabulary

The debate over whether reading comprehension has no subskills or many subskills is a long standing one among reading professionals. A series of articles that appeared in the Reading Research Quarterly between 1968 and 1973 on this topic argued various points of view (Davis, 1968, 1972; Spearitt, 1972; Thorndike, 1973). However, they all agreed that if there are any subskills word knowledge is one of them, and if there are no subskills it is because word knowledge and reasoning in reading are indistinguishable.

In a well planned and carefully executed study, Davis (1968) analyzed a large data set around eight hypothesized comprehension subskills. Items were constructed so that each selection had only one corresponding question. As a result, all items were independent of each other. Davis'

original analysis yielded five unique skills: (a) word knowledge, (b) finding answers to questions that were asked either explicitly or in paraphrase, (c) inferencing, (d) recognizing the writer's purpose, and (e) following text structure. A later factor analytic study (Davis, 1972) revealed four skills: (a) word knowledge, (b) obtaining meaning from context, (c) finding answers to questions that were asked explicity or in paraphrase and weaving these ideas together in content, and (d) drawing inferences. Spearitt (1972) reanalyzed the same data set and obtained four skills: (a) word knowledge, (b) drawing inferences, (c) recognizing a writer's purpose, and (d) following text structure.

In all of these studies, word knowledge was the most robust subskill. In fact, Spearitt (1972) showed that if the effects of word knowledge were removed from the analysis, all other potential skills would reduce to one: reasoning. Thorndike (1973) took that stance one step further. Working with Davis' (1968) original data set, Thorndike initially found three subskills. However, the first one to emerge accounted for 93% of the variance. Ultimately, he concluded that there was no substantive distinction between word knowledge and reasoning in reading because very little difference was found in his analysis between word knowledge and paragraph comprehension. After analyzing the Davis (1968, 1972) and Spearitt (1972)

studies, Rosenshine (1980) concluded that "one sees that different analyses yield different unique skills, and only one skill was consistent across the three analyses: remembering word meanings" (Rosenshine, 1980, p. 543).

Thus it is possible that vocabulary is the most critical skill that can be taught in relation to reading comprehension. Therefore, teachers should be aware of the demands of the vocabulary in a given story, teach that vocabulary, and assess students' grasp of it both prior to and following reading. For this reason, oral retellings cannot adequately be assessed without reference to the student's understanding and use of the appropriate vocabulary.

Background Knowledge

The argument for asking teachers to assess their students' background knowledge with respect to the demands of a given story is almost self-evident. The entire focus of schema theory, reviewed previously in this chapter, is that one's background knowledge is organized into a series of cognitive structures, or schemata, which are responsible for making sense of the world and guiding the process of comprehension in reading (Rumelhart, 1980). Without sufficient background knowledge, the needed schema would not be available and comprehension would be difficult at best. Furthermore, Tierney and Cunningham have concluded that "intervention research has supported the existence of a

causal relationship between background knowledge and comprehension" (1984, p. 612).

Several studies, such as that reported by Kintsch and Greene (1978), have demonstrated the difficulties caused by insufficient schemata or background knowledge. Using stories selected from a collection of Alaskan mythology, the authors demonstrated that even sophisticated readers have difficulty summarizing stories for which they lack the correct schemata. Thus, stories that are typical and easily comprehended within one culture are seen as bizarre and are poorly comprehended by readers from another culture (Kintsch & Greene, 1978, p. 4).

While insufficient background knowledge has been shown to have detrimental effects on comprehension, enriching students' prior knowledge has been shown to have positive effects. Graves, Cooke, and LaBerge (1983), for example, have found that incorporating strategies intended to increase background knowledge into previews of short stories significantly increased students' comprehension.

The anecdotal accounts noted in the "Intrusions" section of this review illustrate the positive effects of a strong background knowledge. The boy who knew how to saddle and ride a horse could identify with the difficulty of the heroine's task in the story. The second boy, who compared the heroine to Paul Revere, already had a schema that included events, goals, outcomes, and resolutions similar to

those of the new story. Comprehension was enhanced as a result. Therefore, an adequate assessment of a student's ability to understand a specific story should include some reference to the fit between the student's prior knowledge and the demands of the new passage.

Summary

The purpose of the six preceding sections of this review has been to demonstrate the importance of each of those constructs either to efficient reading comprehension or to its diagnosis. It has been argued that a person's schemata are constructed from his or her own background knowledge. These schemata rely heavily on inferencing skills to fill uninstantiated slots. One particular schema, a schema for stories, gives the reader a sense of the appropriate sequence of events within a story and a method of judging what is important, what should be retained in memory, and what should be reproduced during recall. The analysis of oral recalls is greatly enhanced if attention is paid to any intrusions made by the student because intrusions can reveal much to the teacher about the nature and quality of the student's comprehension. Furthermore, it seems safe to conclude that comprehension relies heavily on the student's knowledge of words. Finally, students' ability to monitor their own comprehension has been shown to contribute to its eventual quality and completeness. Since all of these items play a crucial role in comprehension or

its diagnosis, each of them appears on the retelling checklist in the form of an open-ended question.

Summary of Chapter Two

The present research project is intended to test the effectiveness of using oral retellings as an informal technique for diagnosing reading comprehension of basal stories. This chapter has reviewed the literature pertaining to three aspects of that technique: (a) its theoretical foundation, (b) the successes and failures reported by others who have investigated the technique, and (c) the inclusion of six open-ended questions on the checklist which direct teachers' attention to variables which affect reading comprehension and its assessment.

Based on this review, it appears that schema theory offers a highly efficient and robust explanation of certain phenomena involved in reading comprehension. The existence of a schema for stories seems to be a particularly consistent finding by a number of researchers. That schema involves a number of predictions which have been verified concerning the nature of students' comprehension of stories. One technique that has a long history as a research tool in the examination of comprehension is the use of oral retellings. Recently, there has been considerable interest, and some reported success, in the use of retellings as an informal method for the diagnosis of comprehension in the classroom. Coupling retellings with schema theory and the

proposed structure of stories seems to be a particulary promising approach to that diagnosis. The addition to that approach of some open-ended questions is intended to help teachers become aware of several important variables that affect comprehension.

The Carroll Oral Retelling Checklist found in Appendix A of this document was constructed to reflect schema theory and to capitalize on students' sense of the structure of stories. It also draws on previous research into the use of retellings in the classroom. If the technique proves to be an efficient means of obtaining an informal diagnosis, then this project will provide evidence to support the claims of Mitchell et al. (1984). If the technique proves to be fruitless, further research will be required before any conclusions can be drawn.

CHAPTER THREE
PROCEDURES

Introduction

The specific idea for this project arose from the content of Pre-Convention Institute #17 at the 1984 Annual Convention of the International Reading Association. The presenters of Institute #17 (Mitchell, Clark, Grant, Irwin, Marshall, Mason, & Leys, 1984) argued that oral retellings can be used to obtain an informal diagnosis of student comprehension. Several of them (J. Grant, 1984; Marshall, 1984b) specifically suggested the use of story grammar research to provide a framework for developing an assessment instrument and offered examples of instruments they had developed. Their efforts, reviewed in Chapter Two of this paper, provided the impetus for this research project.

This project was comprised of four separate investigations. The first two examined the validity and reliability of the Carroll Oral Retelling Checklist (CORC). The other two investigations examined the impact of repeated use of the oral retelling technique on students' oral retelling behaviors and on teachers' ability to make informal assessments of students' comprehension strengths and weaknesses. These four investigations address the following hypotheses:

1. There will be no statistically significant correlation between subjects' standardized reading comprehension test scores and scores assigned to them through use of the oral retelling technique.

2. There will be no statistically significant correlation among the scores assigned to 10 different retellings of a single story by five trained users of the oral retelling checklist.

3. There will be no statistically significant difference between the final retellings of students who have had repeated exposure to the oral retelling technique and retellings of the same story done by students who have not been exposed to the technique.

4. There will be no statistically significant difference between the quality of the informal reading diagnostic assessments developed by teachers who have used the oral retelling technique and by those who have not, as judged by a panel of experts in the field of reading.

Development of the Checklist

In order to address these four research hypotheses, and in order to ensure uniformity across all four investigations in this study, a decision had to be made either to adopt one of the checklists previously published, adopt one and modify it in some way for use in this study, or design a new one specifically for this study. It was decided to develop a new checklist because the answer to the fourth research

question involved teachers' diagnostic judgments. Such impressions must go beyond what students do or do not recall to variables that affect the quality of that recall (Johnston, 1983). Therefore, the checklist includes some of those variables.

The "product data" section of the Carroll Oral Retelling Checklist developed for this study (Appendix A) is grounded in a review of story grammar research. The categories listed on the checklist are similar to those most frequently cited in the literature as exemplifying an adequate story grammar (See Chapter Two). The terminology used on this portion of the checklist, however, is slightly different than that which is usually found in the literature. This was done at the suggestions of teachers who were somewhat uncomfortable with categories such as "internal reactions" or "goal path," and who preferred the inclusion of "characters" as a separate category.

The "process data" portion of the checklist reflects several key variables that influence student comprehension. These varaibles were suggested by Mitchell et al. (1984) who contended that they could be assessed through the analysis of oral retellings. Chapter Two of this document contains a brief review of some of the pertinent literature which established the importance of these variables to efficient comprehension.

After the checklist was developed, it was piloted by the researcher with a small group of fifth graders, and some weaknesses were readily apparent. The original checklist was not divided into the two categories of product and process data. It called for the assessment of a student's metacognitive awareness using the same one-to-five scale that was used to assess the products of comprehension. Measuring process variables in this way is unrealistic; yet the information gained by directing teachers' attention toward an awareness of them seemed valuable. As a result, the checklist was changed so that the items which are process oriented, and less easily assessed, are listed as open-ended questions. This led to a logical division of the checklist into product and process sections.

The Training Sessions

Once the checklist was completed, the next step was to train the educators who participated in this project in the use of the oral retelling technique. Two different types of training sessions were required for this study. The first, which lasted about 30 minutes, provided the orientation and preparation for the teachers in the control group of Experiments 3 and 4. The second, which lasted about 75 minutes, provided the training for the public school teachers and doctoral students who were required to use the checklist.

The Control Group

The six teachers who participated in this study as the control group met with the researcher for one 30-minute session near the end of the experimental period. At that time, their role in the experiment was explained to them. The students who served as the control group for Experiment 3 were not identified until a short time before the administration of the posttest in order to avoid the possibility of preferential treatment affecting the results of the study.

The Experimental Group

Separate training sessions were held for the public school teachers and doctoral students who participated in Experiments 2, 3, and 4. However, all training sessions followed the format described below.

Each session began with the distribution of the checklist to all participants. The researcher explained the directions for administering the checklist, defined all technical terms, and answered all questions until, in the researcher's opinion, all participants understood both the procedures and the focus of each of the items.

Once the checklist was thoroughly discussed, the session continued with the presentation of a tape-recorded retelling of a story done by a fifth-grade student who participated in Experiment 1. Each participant was given a copy of the printed form of the story and a copy of the

researcher's scoring of the retelling. The researcher then modeled the scoring process and explained the rationale for each scoring decision. Pertinent portions of the recording were replayed to illustrate key points. When all participants' questions were answered, this step was repeated, using a second retelling of the same story to illustrate additional points not emphasized in the first retelling.

The third step in the training procedure began with the division of the participants into small groups. A third retelling of the same story was presented and each group was asked to produce a collaborative scoring, discussing each item as they scored it. When the process was completed, the whole group reassembled and compared their efforts. After the discussion was completed, this step was repeated with a new retelling.

The fourth step called for individual scorings of another retelling of the same story followed by a whole-group discussion. Again, the step was repeated after a discussion of the scorings.

Finally, participants were presented with a printed copy of a new story and a tape-recorded retelling. They were asked to score the retelling without conferring with anyone else. When all of the scorings were completed, participants compared their results and discussed them as needed.

The training session for the public school teachers concluded with a full description of the research project, each teacher's rights and responsibilities, their students' rights and responsibilities, and the procedures for obtaining the informed consent of their participating students and the students' parent(s).

In summation, the training sessions for those who were to use the CORC followed this basic pattern: (a) a discussion of the procedures for administering the checklist and of the items included on the form; (b) two instances of the modeling of the use of the checklist by the researcher, with a discussion following each; (c) two instances of group scorings, with ensuing discussions; (d) two instances of individual scorings, with discussions, and (e) a single scoring of a retelling of a new story.

Experiment 1

This experiment was designed to investigate the degree to which scores given students on the ORC would correlate with their scores on the Intermediate I level of the Metropolian Achievement Test (Prescott, Balow, Hogan, & Farr, 1978). It utilized a Pearson product moment correlation between ratings assigned to students' retellings of stories and their standardized reading comprehension scores. The resultant correlations produced an estimate of the validity of the checklist as a reading comprehension diagnostic device.

Sample

The sample population for this experiment included 30 fifth graders from a K-12 school in north central Florida. They comprised one intact class of students who had been randomly assigned to that class at the beginning of the school term. These students have a wide range of ability levels and are from mixed socio-economic, racial, and ethnic backgrounds. However, only those students who have English as their primary language participated in this study.

Materials

The story used in this study is an edited version of "Sybil's Midnight Ride" by Eva Grant (Grant, 1980). The story was slightly altered by the researcher in order to lower its readability level. In its edited form, it has a readability level of 5.0 according to the Fry Readability Formula (Fry, 1977).

Instrumentation

Reading comprehension scores for the subjects were taken from the Intermediate 1 level of the Metropolitan Achievement Test (Prescott et al., 1978) administered in April, 1984. Oral retelling scores were obtained through the use of the CORC developed for use in this research project (Appendix A).

Methodology

All students in the participating class who have English as their primary language were randomly assigned a

number. This number was used for identification purposes. Then, the first student was asked to meet with the researcher in a location away from other students and teachers. The procedures that were to be followed were explained to the subjects and they were asked to read the story. Immediately afterwards, they were asked to retell the story as accurately as possible to the researcher who recorded the retellings on tape. Specifically, each student was asked: "Now that you have finished reading the story, will you please retell it to me in your own words as if I had never heard it before?" At the end of the retelling the subjects were asked: "Is there anything else about the story that you can remember, or is there any comment that you would like to make?" Responses to these questions were also recorded on tape and included in the analysis. At a later time, the researcher reviewed and scored each of the retellings using the CORC (assigning five points to the "All" category, four points to the "Much," etc.).

Subjects for this experiment were identified through the use of a coding system in order to protect their right to privacy. The teacher of the subjects kept a list of the participating students, the order in which they were sent to the researcher, and their standardized reading comprehension test scores. Only after all of the retellings had been assessed by the researcher was the list of students'

standardized test scores supplied for data analysis, again using the coding system rather than names.

Data Analysis

The two lists of scores (the oral retelling scores and the standardized test scores) were subjected to a Pearson product moment correlation. This determined the degree of correlation between the standardized test score and the retelling scores. The resultant correlation yielded a measure of the degree to which a trained user of the checklist is able to differentiate between good and poor readers as they are measured by a standardized reading comprehension test and, therefore, a measure of the validity of the instrument.

Experiment 2

While the first part of this study focused on the validity of the CORC, the second experiment was designed to provide an estimate of its inter-rater reliability. This experiment employed a series of Pearson product moment correlations among the various scores assigned by trained users of the checklist to 10 separate retellings of a single story by fifth-grade students.

Sample

The subjects utilized in this study included five educators who had been trained in the use of the CORC as described above. Among them were two teachers from the treatment group of Experiments 3 and 4 of this research

project, and three doctoral students in reading. The teachers were drawn from the Chapter I reading staff of a county school system in north central Florida. Both of them are certified by the State of Florida to teach in the elementary grades and have a minimum of 11 years teaching reading at the elementary level. Both subjects are female and both were in their mid-30s. The doctoral students included one male and two females, all of whom have had experience teaching elementary and/or middle school reading. Their ages range from 27 to 35, their average length of experience teaching reading was seven years.

Materials

Ten tape-recorded oral retellings of "Sybil's Midnight Ride" (Grant, 1980), obtained during Experiment 1, were used in this study.

Instrumentation

The CORC was the only instrument required for Experiment 2.

Methodology

All subjects in Experiment 2 were trained in the use of the CORC employing the procedures described above. When the training was completed, each participant independently listened to and scored the 10 retellings. Each retelling was preceded by a code number, and the participants were asked to use that code number on their completed checklists.

Data Anaysis

Pearson product moment correlations were calculated for all subjects on each retelling. The resultant correlations produced an estimate of the level of inter-rater reliabililty achieved on the checklist after the users were trained as described above.

Experiment 3

This experiment was designed to investigate whether repeated exposure to the retelling technique would result in more complete retellings by students. It utilized an analysis of variance between experimental and control group retellings of a single story to test for significant differences in the quality of the retellings.

Sample

This study utilized two different sample populations. The first included the participating teachers, and the second included the participating students.

The teachers who participated in this experiment included 12 members of the Chapter I teaching staff of a county school system in north central Florida. All subjects were certified by the State of Florida in two areas: elementary education and reading. Subjects' ages ranged from 28 to 45 with an average of 10 years' experience teaching reading at the elementary school level. All subjects volunteered to participate in this project, and

were randomly assigned either to the experimental or control condition.

The students for this study were randomly selected from the classes of the twelve participating teachers. Between four and six subjects were chosen at random from each teacher's class. This resulted in a total of 60 subjects, 30 in each group.

All students participating in this study qualified for inclusion in the Chapter I program of the participating school district located in north central Florida. In order to qualify, a student must score a minimum of two years below grade level in reading or mathematics on the Metropolitan Achievement Test (Prescott et al., 1978). Only those students who scored more than two years below grade level on the reading portion of that test were eligible for this study. The students represented a wide range of socio-economic, racial, and ethnic backgrounds, but all had English as their primary language. Their ages ranged from 10 to 13 years. There were more boys (36) than girls (24), presumably due to the traditional make-up of basic skills classes at the elementary school level.

Materials

All students and teachers participating in this portion of the research project used the Ginn Reading Program (Clymer, Venezky, Johnson, & Pearson, 1982) for their normal daily reading lessons. Because of the nature of the basic

skills classes, several levels of the Ginn system were in use in each of the classrooms. The only important difference in materials used in the experimental and control classes was the oral retelling checklist.

The posttest for this study involved the oral retelling of the story "Penny's Good Fortune" (Eller & Hester, 1980) from Level 10 of the Laidlaw Reading Program. The readability level of the story was 3.5 according to the Fry Readability Formula (Fry, 1977).

Instrumentation

Again, the only instrument required for this experiment was the CORC.

Design of the Experiment

Experiment 3 utilized a two-group, randomized-subjects, posttest-only design. It can be represented graphically by the following:

	Group	Ind. Variable	Posttest
(R)	E	X	Y
(R)	C	---	Y

In the above diagram, (R) represents random assignment of teachers to experimental (E) and control (C) conditions and random selection of students within the teachers' intact classes. As such, this is actually a quasi-experimental design. The use of the oral retelling procedure is represented by (X). The posttest comparing the retellings of the two groups is represented by (Y).

Methodology

In August, 1984, the researcher spoke to a meeting of all Chapter I teachers in the participating school system. The project was described in general terms and teachers were asked to volunteer to participate. The director of the county's Chapter I program expressed support for the project. Subsequently, 12 teachers volunteered. They were randomly assigned to the experimental and control conditions, six to each. The teachers in the experimental group were trained in the use of the oral retelling technique as described above.

During the experimental period, which lasted from mid February, 1985, until mid May, 1985, teachers in the treatment group used the oral retelling technique once each week in each reading group that had a participating student. Since no group had only one participating student, each subject had an opportunity to be the storyteller some weeks and to listen to a storyteller during other weeks. On average, each subject served as the storyteller five or six times during the experiment.

The oral retelling technique used by the teachers with their students was the same as used by the researcher in Experiment 1. Before silent reading began a storyteller was identified who, after completing the reading, was asked to retell the story. Other students in the reading group were cautioned not to prompt the storyteller. In each case the

teacher used the following directions: "Now that you have finished reading the story, will you please retell it to us in your own words as if we had never heard the story before?" After the retelling was completed, the teacher added: "Is there anything else about the story that you can remember or any comment that you would like to make?" The entire retelling and the response to the follow-up question were recorded on tape.

Each subject in the treatment group was assigned a single audio-tape for use throughout the experiment. As a result, all of student X's retellings are on a single audio-tape, and are in the exact sequence in which they were told; that is, the earlier experiences are located toward the beginning of the tape, and the later ones toward the end. Each tape had a code number to identify it for the teacher. However, the researcher had no indication of the identity of any of the subjects.

Two weeks before the administration of the posttest for this study, the students for the control group were randomly identified. During the experimental period, these students received their reading instruction using the same basal reading series used by the experimental group.

There were at least two readily apparent threats to the internal validity of this experiment. One possible threat to the integrity of the information obtained during the posttest was the presence of the tape-recorder itself. The

students in the experimental group had been using the tape-recorder for nearly three months prior to the posttest and were accustomed to its presence. In order to minimize this potential threat to internal validity, all control group subjects had two opportunities to use a tape recorder during the two weeks prior to the posttest. Specifically, they engaged in at least two language-experience stories. This is a commonly used technique wherein students are asked to tell a story about something that is meaningful to them. Each story was recorded on tape and used as a basis for a reading lesson.

The second threat to internal validity involved the process of telling stories. It was possible that the subjects in the experimental group would perform better on the posttest simply because they had experience telling stories, as opposed to better comprehension. The language experience approach used during the last two weeks prior to the posttest gave the control subjects practice in oral storytelling and minimized the impact of storytelling itself on the outcome of the experiment.

Three critical underlying assumptions in this investigation must be addressed. The first is that teachers in the experimental group would not give their students direct instruction in how to retell a story. That practice would yield better retelling performances but would negate any implications that might be drawn about the effectiveness

of the technique as a means of diagnosing reading deficiencies. The second assumption was that, once teachers obtain diagnostic information through the use of the checklist, they would use that information to remediate any perceived weaknesses. The third assumption was that remediating comprehension weaknesses would lead to improved retellings. If the first two assumptions could be verified, the statistical results of this investigation would have implications as to the validity of the third.

In order to verify those two assumptions, qualitative data were collected during the course of this investigation. The researcher periodically visited the classroom of each of the teachers in the experimental group. Reading lessons, including the use of the oral retelling technique, were observed and anecdotal accounts described in the researcher's field notes. Teachers were asked about their impressions of the effectiveness of the technique, the kinds of information it revealed to them, and how they used that information in their teaching. A combination of teachers' reports and the observations of the researcher added some evidence as to validity of the first two assumptions and shed some light on the results of the posttest.

The posttest for this experiment consisted of a single retelling of the story "Penny's Good Fortune" (Eller & Hester, 1980). All subjects in both the treatment and control groups were asked to read the story silently and to

retell it using the same technique as described above. The retellings were recorded on tape for later analysis. The posttest retellings were coded so that they contained no indication of whether they came from the control or treatment group.

In summation, both the experimental and control groups received their reading instruction using the Ginn Reading Program (Clymer et al., 1982) basal series. The experimental group, however, also used the oral retelling technique described in Experiment 1 once each week with each participating reading group, and approximately five or six times with each subject. In the control group, use of the language experience technique served a dual purpose. It gave each subject an opportunity to use and become familiar with the tape-recorder, and it gave each subject some storytelling experience. Removal of these threats to internal validity allowed more accurate generalizations based on the results of the posttest. A posttest was administered to all subjects and the results scored by a trained user of the checklist who had no indication of the origin of each retelling.

Data Analysis

An analysis of variance was computed to test for significant differences in the scores assigned to the retellings of the treatment and control groups. Qualitative

data were also collected and analyzed throughout the experimental period.

Experiment 4

The final study in this research project was designed to investigate the effects of the use of the oral retelling technique on teachers' ability to listen to a retelling and make an informal diagnosis of the quality of comprehension evident within it. It utilized a z-test for independent proportions to test for significant differences between the proportion of treatment and control teachers' assessments identified by a panel of experts in reading as being "rich and usable" as defined in this study.

Sample

The population for this study was comprised of the same 12 teachers who participated in Experiment 3. Six of them (the experimental group) were trained in the use of the oral retelling technique and employed it with their own students (as described in Experiment 3) for a period of 10 to 12 weeks. The other six teachers who volunteered for this project comprised the control group.

The panel of experts used in this study included three members of the faculty of a public university in Florida. They all possess doctoral degrees with a concentration in reading and all have extensive experience teaching graduate level reading courses.

Materials

A single retelling of the story "Brush with White Death" by Francis Shell (1977) collected during a pilot study of this investigation was used in this experiment. All teachers were supplied with a printed copy of the story and an audio-tape with a single retelling of that story. Therefore, all teachers heard the same retelling.

Instrumentation

It would have been both unfair and invalid to ask the six untrained teachers to use the ORC as an assessment instrument. Therefore, an alternative assessment method had to be found that could be used by all twelve subjects. Thus, for the purpose of this study, all subjects were asked to write a narrative assessment of the quality of comprehension exhibited in the retelling supplied to them and to speculate as to the factors that might have contributed to any perceived weaknesses. They were also asked to speculate on the remedial steps that should be taken, if any.

Design of the Study

This experiment utilized a two-group, randomized-subjects, posttest-only design similar to that described in Experiment 3. In the diagram:

	Group	Treatment	Posttest
(R)	E	X	Y
(R)	C	---	Y

(R) represents random assignment of the teachers to

experimental (E) and control (C) conditions. The use of the oral retelling technique is indicated by (X). The posttest, comparing the proportion of experimental and control teachers' assessments identified as being "rich and usable," is indicated by (Y).

Methodology

The use of the oral retelling technique by the experimental group was identical to that used in the experiments described above. The control group had no exposure to the technique.

At the conclusion of the experimental period, all 12 teachers were given a copy of the posttest tape, a printed copy of the story being retold, and a booklet to be used for their assessments. After reading the story and listening to the retelling, all teachers wrote out a narrative evaluation of their perception of the quality of the students' comprehension of the story. Teachers were asked to include as much information as they could about the accuracy of the retelling, the possible causes of any perceived weaknesses, and possible remedial steps that could be taken.

The narrative assessments were collected and randomly assigned a number from 1 to 12, to be used for identification purposes. The assessments were then given to a panel of three experts in reading who judged them using a holistic approach commonly used in assessing student essays. The experts were asked to read the story, listen to the

retelling, and make a series of judgments on the overall quality of each of the narratives. The narratives were judged on the basis of accuracy, thoroughness, and usability in planning a remedial program. After judging all narratives, the experts separated them into two groups in order to identify the better ones.

Data Analysis

The data analysis procedure for this study involved the use of three z-tests for independent proportions, one for the assessments made by each expert.

Summary

In summation, this research project was comprised of five steps. The first included the development, testing, and revision of the instrument. The second step involved the training of a group of educators in the use of the oral retelling diagnostic technique. The third step included two preliminary investigations that addressed the reliability and validity of the instrument. The fourth step included an investigation into the impact of repeated exposure to the oral retelling technique on the retelling behaviors of fourth- and fifth-grade students. This involved the implementation of the technique by six Chapter I teachers in their own classrooms for a period of approximately 10 to 12 weeks, and the collection of both quantitative and qualitative data. The last step encompassed an

investigation into the impact of the use of the technique on teachers' diagnostic insights.

CHAPTER FOUR
RESULTS OF THE INVESTIGATION

Introduction

The preceding chapters have described an investigation of the use of oral retellings as an informal means of diagnosing reading comprehension. The Carroll Oral Retelling Checklist (CORC) was developed for use in this investigation which incorporated an approach to assessing comprehension based on research into the structure of stories. The investigation was comprised of four separate studies. The first correlated the oral retelling scores of 30 fifth-grade pupils with reading comprehension scores obtained on the Intermediate 1 level of the Metropolitan Achievement Test (Prescott, Balow, Hogan, & Farr, 1978). The intent was to provide an estimate of the procedure's validity with regard to one widely respected standardized test. The second study correlated the scores given to 10 retellings of a single story by five educators who were trained in the use of the technique. The resulting correlations were intended as an estimate of the inter-rater reliability of the procedure.

The final studies involved the use of the retelling checklist and procedures in six Chapter I basic skills classes at the fourth- and fifth-grade levels. Six

teachers, working with a total of 30 students, used the technique as a supplement to their normal reading lessons for a 12-week period. In order to test the impact of the procedure on students' comprehension, the third study compared the experimental and control groups on a retelling of a single story. The final study compared the teachers in the experimental and control groups on their ability to diagnose the quality of a child's comprehension based on a retelling of a story. This final measure was intended to assess the impact of repeated use of the technique on teachers' diagnostic insights.

This chapter presents the data collected during the investigation and the results of all statistical analyses. Chapter Five discusses the implications and conclusions that may be drawn from this research.

Experiment 1

Experiment 1 was designed to investigate the first hypothesis that there would be no statistically significant correlation between subjects' standardized reading comprehension test scores and scores assigned to them through the use of the CORC. To test this hypothesis, 30 fifth-grade students read and retold the story "Sybil's Midnight Ride" (Grant, 1980). Their retellings were assessed using the CORC and the results were correlated with the students' reading comprehension scores from the

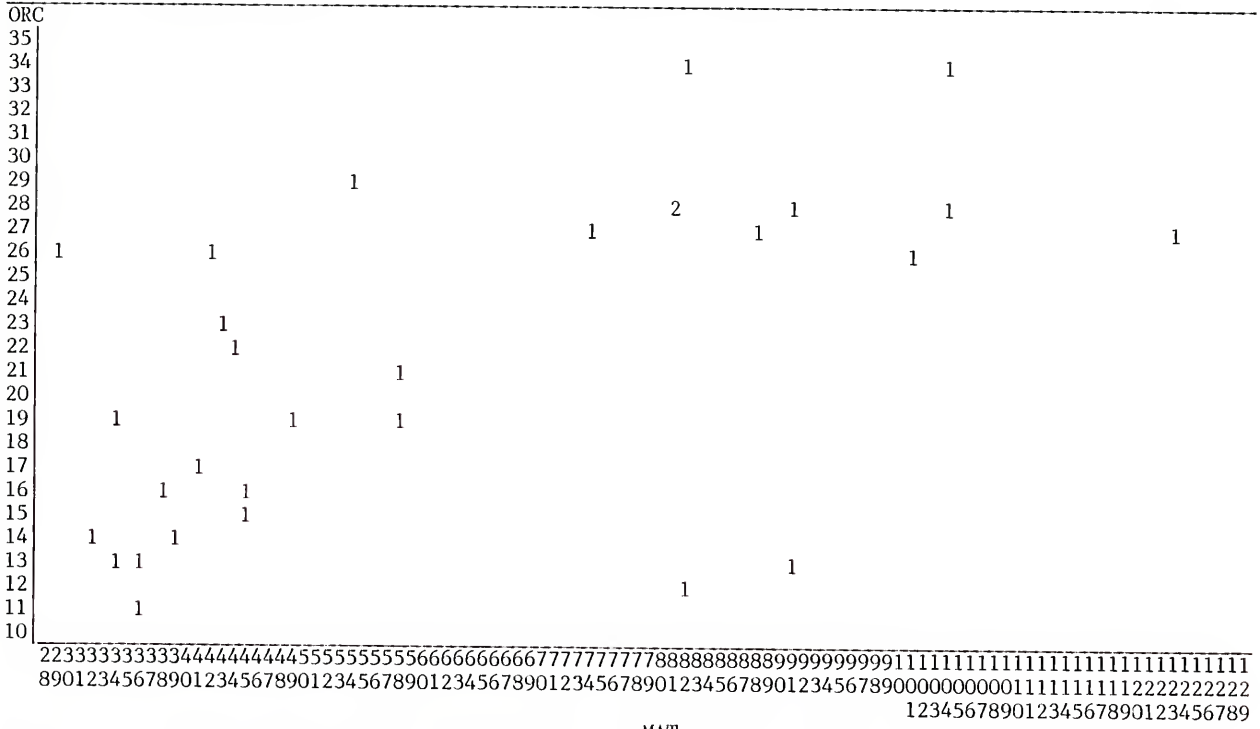
Table 1: Scores of fifth-grade students on the Carroll Oral Retelling Checklist (CORC) and the Intermediate 1 Level of the Metropolitan Achievement Test (MAT).

Student	CORC	MAT	Student	CORC	MAT
1	11	3.6	16	14	3.2
2	12	8.2	17	15	4.5
3	19	3.4	18	19	5.8
4	21	5.8	19	26	4.2
5	27	7.4	20	19	4.9
6	13	9.1	21	16	4.5
7	29	5.4	22	17	4.1
8	27	12.3	23	26	2.9
9	27	8.8	24	26	10.1
10	28	8.1	25	28	9.1
11	13	3.4	26	14	3.9
12	13	3.6	27	34	8.2
13	28	8.1	28	23	4.3
14	34	10.4	29	28	10.4
15	16	3.8	30	22	4.4

Metropolitan Acievement Test (Prescott et al., 1978). The scores from these two tests are listed in Table 1.

After the data collection was completed, a scatter plot was prepared and is reproduced in Figure 1. It was immediately evident that four points differed markedly from the rest. Three subjects (Students 2, 6, and 8) had MAT scores that were much higher than would have been indicated by their performance on the CORC. A fourth subject (Student 23) scored much lower on the MAT than would have been predicted. These four cases were investigated more thoroughly in an attempt to understand the reasons for the disparities and to shed more light on the apparent inability of the CORC to provide an adequate assessment of the students' comprehension.

Conversations with the teacher of all four subjects produced some interesting information that will be discussed



Note: Read all MAT scores as if the final digit was preceeded by a decimal point; that is, read 101 as equal to a grade equivalent of 10.1.
Digits in the scatter plot indicate the number of subjects who scored at that point.

Figure 1: Scatter plot comparing student performance on the ORC and MAT tests.

in more detail in the next chapter. However, for purposes of the analysis of the data collected in Experiment 1, three points are of particular interest. First, the teacher characterized students 2 and 6 as very capable pupils whose MAT scores accurately reflected their classroom performance. She added that both students were shy and generally became very quiet around visitors to the classroom. Both students were tested by the researcher on his first day spent in that classroom. Therefore, their shyness may have inhibited them from fully reporting what they remembered from the story. Subsequent retellings performed by the same students during another part of this investigation seemed to confirm this possibility. Both students obtained retelling scores that were much higher than their original scores.

Student 23 presented a different kind of problem. Her MAT comprehension score was reported as 2.9 but she did a very good retelling when tested by the researcher. The teacher discounted the MAT score and produced the student's prior testing history. For every other administration of the MAT the student had scored on or slightly above grade level in reading comprehension. Therefore, it seems possible that this particular MAT score may be invalid.

The final case (Student 8) was the most confusing. The student's history revealed reading scores that were consistently above grade level, though never to the extent achieved with this result of a 12.3 reading comprehension

grade equivalent on the MAT. Furthermore, the teacher had administered a Slosson Oral Reading Test (SORT) (Slosson, 1963) to the student and obtained a reading score of 6.8. This score was still 1.5 years ahead of grade placement but 5.5 years below the MAT score. The student's CORC score of 27 placed among the top third of the class and would have been in closer agreement with the SORT score than the MAT. However, no logical reason for the disparity could be offered by the teacher or discovered by the researcher.

With the above information in mind, it was decided to perform and report two Pearson product moment correlations. The first included all 30 data points and the second excluded those of students 2, 6, and 23. The scores of Student 8 were retained in the second analysis because no evidence could be produced to explain the disparity between the CORC and the MAT score.

The initial correlation, including the entire data set, indicated that a significant direct relationship existed between the subjects' performance on the ORC and their standardized reading comprehension scores obtained on the MAT ($r = .5779$, $p < .01$). The second correlation ($r = .7866$, $p < .005$) was considerably stronger since the three disparate scores were eliminated. Based on these results it would appear that a significant relationship between the two tests does exist and that the true strength of that

relationship may lie between .57 and .78. Therefore, the first hypothesis of this investigation was rejected.

Experiment 2

The second hypothesis was that there would be no statistically significant correlation among the scores assigned to 10 different retellings of a single story by five trained users of the CORC. Therefore, an experiment was designed to provide an estimate of the inter-rater reliability of the oral retelling assessment procedure.

Five raters independently assessed 10 retellings of a single story using the CORC. Table 2 displays the raw scores given by each of the raters to the retellings. The rankings of the 10 students based on the CORC scores of each rater are displayed in Table 3. Pearson product moment correlations were calculated for all possible pairs of raters and are reported in Table 4. The lowest correlation was .8611 ($p < .005$) and the highest was .9761 ($p < .005$) with a mean of .9187 ($p < .005$).

These results tend to indicate that a high degree of inter-rater reliability exists when trained users of the ORC employ it to assess the quality of student comprehension revealed in their retellings of stories. Therefore, the second hypothesis of this investigation was rejected.

Experiment 3

The third experiment examined the impact of participation in retelling activities on students' retelling

Table 2: CORC scores given to 10 retellings by five raters in Experiment 2.

Retelling	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
1	16	18	14	25	21
2	25	33	27	30	27
3	25	33	24	31	28
4	11	14	12	16	13
5	17	21	12	26	14
6	16	22	18	24	15
7	25	36	29	33	26
8	12	13	14	15	14
9	32	39	34	37	34
10	15	16	16	24	22

Table 3: Rankings of each retelling by five raters in Experiment 2.

Retelling	Rank by Rater 1	Rank by Rater 2	Rank by Rater 3	Rank by Rater 4	Rank by Rater 5
9	1	1	1	1	1
7	3	2	2	2	4
2	3	3.5	3.5	4	3
3	3	3.5	3.5	3	2
5	5	6	9.5	5	8.5
6	6.5	5	5	7.5	7
1	6.5	7	7.5	6	6
10	8	8	6	7.5	5
8	9	10	7.5	10	8.5
4	10	9	9.5	9	10

Table 4: Inter-rater correlation matrix for Experiment 2.

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
Rater 1	----	.976	.948	.952	.924
Rater 2		----	.947	.945	.861
Rater 3			----	.862	.903
Rater 4				----	.881
Rater 5					----

behaviors. Specifically, it tested the hypothesis that there would be no statistically significant difference between the quality of the final retellings of students who had been repeatedly exposed to the oral retelling technique and retellings of the same story done by students who had not been exposed to the technique. To probe this hypothesis, teachers in an experimental group used the retelling procedure in their classrooms with randomly selected students for a period of 10 to 12 weeks. Then both experimental- and control-group students read and retold a single story. The retellings were assessed using the CORC and the results subjected to a one-way analysis of variance. The raw scores given to the retellings of the two groups are reported in Table 5 and the results of the analysis are reported in Table 6.

On average, students in the experimental group scored 3.933 points higher on the posttest than students in the control group. The results of the ANOVA indicated that this difference in performance was significant at the .05 level ($F = 5.11$, $df = 1/58$, $p < .05$). These results suggest that participation in the experiment was of some benefit to the students and led to their performing significantly better on the posttest than their counterparts in the control group. Furthermore, because the control-group students had practice in telling stories and in using the tape recorder and took the posttest in familiar environs, it would appear that

Table 5: Results of the posttest of Experiment 3 comparing treatment and control students on a single retelling.

Treatment	Control	Treatment	Control
12	12	23	27
26	21	29	19
21	11	33	21
27	19	32	12
12	12	26	11
14	30	32	12
8	21	29	16
12	15	25	18
30	18	21	20
19	12	31	25
19	21	29	21
28	29	17	31
31	19	31	18
27	21	17	29
14	25	31	22

Table 6: Analysis of variance summary table for the post-test performance of the experimental and control groups in Experiment 3.

Source	Sum of Squares	df	Mean Square	F
Between	232.0667	1	232.0667	5.1126*
Within	2632.6666	58	45.3908	
Total	2864.7333	59		

* $p < .05$

something other than practice in telling stories contributed to the better performance of the experimental group. Based on these findings, the third hypothesis was rejected.

Experiment 4

The final experiment in this investigation examined the impact of repeated use of the oral retelling assessment procedure on the diagnostic insights of the experimental-group teachers. It addressed the fourth hypothesis of this investigation that there would be no statistically significant difference between the quality of the diagnostic assessments developed by teachers who had used the oral retelling assessment procedure and by those who had not used it.

This hypothesis was tested by comparing the diagnostic assessments that teachers in the experimental and control groups gave to a single retelling of a story. All teachers read the story, listened to its retelling, and then wrote out a narrative stating their assessment of the quality of the student's comprehension. The narratives were judged by three experts in reading working independently of each other. The experts judged the assessments on three criteria: accuracy, thoroughness, and utility in planning a remedial program. A series of z-tests of independent proportions was used to test for significant differences in the quality of the assessments. The results of the

assessments by each judge are listed in Table 7 and the results of the statistical analyses appear in Table 8.

The first reading expert selected six assessments as being clearly better than the rest with only three of them coming from the experimental group. Therefore, no difference existed in either direction. The second expert also selected six assessments for the "better" category. This time, however, five of those selected were from the experimental group. This proved to be significant ($z = 2.286$, $p < .05$). The final expert selected seven "better" assessments and four of them were from experimental-group teachers. This did not represent a significant difference ($z = .597$, $p > .05$). Based on the inconsistency of these results, the fourth hypothesis could not be rejected.

Qualitative Data

The final portion of this chapter summarizes the nine most important findings of the qualitative data collection

Table 7: Diagnostic assessments listed as "clearly better" by three experts in reading.

Expert 1	Expert 2	Expert 3
1 (E)	1 (E)	1 (E)
4 (E)	4 (E)	2 (E)
6 (C)	5 (E)	3 (C)
7 (E)	7 (E)	5 (E)
10 (C)	9 (E)	6 (C)
11 (C)	11 (C)	7 (E)
		11 (C)

Each diagnosis was randomly assigned a number from 1 to 12. Those chosen as being better than the rest are listed above. The letters that follow each number identify them as from the experimental (E) or control (C) group.

Table 8: Z-test summary table for performance on the post-test of Experiment 4 based on the assessments of three experts in reading.

	Number chosen as "better"	Number from exp. group	Z	P
Expert 1	6	3	0	$p < 1.0$
Expert 2	6	5	2.286	$p < .05$
Expert 3	7	4	.597	$p > .10$

that took place during this investigation. The implications of these findings are discussed at greater length in Chapter Five. Three methods were used to collect these data: informal conversations with experimental-group teachers, classroom observations, and structured interviews. The questionnaire used during the structured interviews is included as Appendix B of this document.

Two points are important to a full understanding of the findings outlined below. First, the six teachers in the experimental group were given no specific direction in what to do with the results of their diagnoses. That decision was left entirely to their professional discretion. It was reasoned that if the teachers believed the information to be valuable they would find some way to incorporate it into their daily instruction. The qualitative data collection, therefore, was aimed at discovering whether or not teachers made use of the information obtained during the oral retelling procedure. The second important point is that the teachers were cautioned not to give their students

assistance in doing better retellings. Such assistance would have made it impossible to draw valid conclusions from the posttest results.

The most readily apparent finding to emerge from an analysis of the qualitative data was that all teachers in the experimental group altered their reading comprehension instruction to incorporate a story-structure approach to narrative material. The teachers expressed the belief that the patterns of comprehension revealed to them by the CORC led them to the conclusion that their students needed direct instruction in the structure of stories. Therefore, some of them used story grammar categories to organize discussion of the selection, while others used the categories as the basis for a questioning strategy similar to that of Marshall (1980). These strategies were adopted for use by the teachers with all of their students, not just those who were participating in this study. Furthermore, all of the teachers indicated that they would continue to use both these strategies and the CORC in the future.

Four teachers either reported or were observed using follow-up probes to further diagnose comprehension. These teachers reported that the use of the questions strengthened their diagnosis and often revealed additional details that were recalled by the student but not reported during the retelling.

The third point that emerged from the qualitative data concerned the relative utility of the two portions of the CORC. In the early part of the experimental period, all of the teachers reported that the upper half of the CORC was more beneficial to them than the open-ended questions of the bottom half. However, near the end of that period three teachers reported obtaining some valuable information through their consideration of those questions. For example, one student was apparently unaware that he consistently missed the point of the stories he was reading. Prior to using the retelling procedure this student had been successful at answering numerous questions about the details and events of a story. When asked to retell a story, however, he would include several events and numerous details, but without giving any evidence that he truly understood the story's theme. The teacher realized this while considering the question concerning the child's metacognitive awareness. This new insight helped the teacher address this child's problem during subsequent instruction. Despite this example, however, it appears that the checklist portion of the CORC had greater utility in the eyes of the teachers who used it on a continuing basis.

All six teachers indicated that it took two or three weeks for them to feel comfortable doing the analyses of the retellings. They also noted that it took about the same amount of time for their students to get used to retelling

stories. The greatest benefits and the most accurate diagnoses, in their opinion, came after the first few weeks of using the procedure. This finding is particularly interesting in light of the results of Experiment 4 in which control-group teachers produced diagnoses that were, in the opinion of two out of the three experts in reading, as accurate and thorough as those of the experimental group.

The fifth point was not entirely unexpected, but the extent to which teachers mentioned it was a surprise. All of the teachers indicated that listening to the retelling helped the other children in the reading group focus on the quality of both their own and the storyteller's comprehension. Children would make comments like "that was a good retelling," or "he left out a whole part of the story." Children appeared to be reading the stories more carefully so that they could fill in the events and details that were missing from a retelling. Three teachers specifically reported that the discussion following each retelling was beneficial to the entire reading group and became an integral part of each lesson plan.

All of the teachers in the experimental group were convinced that their students were performing better retellings at the end of the treatment period. The consensus of their opinions attributed that growth to three factors. First, they believed that the students had developed a better sense of the structure of a well-ordered

story. Second, they believed that the retelling procedure forced students to focus on meaning rather than word calling during reading. Third, teachers believed that the instruction and questioning strategies they adopted contributed to improving students' sense of what was important to remember and later to retell from their stories.

The seventh point concerned the utility of the CORC in diagnosing specific weaknesses and planning remediation. All of the teachers could mention at least one child, and often more, for whom specific remedial needs were identified through using the retelling procedure. The teachers added that the information gained by using the ORC was helpful in establishing a remedial plan to meet those needs.

In response to a question during the structured interview, four of the six teachers reported that they would use the technique with all of their students. One said that she would use it with all students but that it would be less effective with those that were habitually non-verbal. The sixth teacher indicated that she had one student who was so non-verbal that using the technique would be more frustrating than helpful.

Finally, all of the teachers reported that the retelling assessment procedure was efficient in that the benefits were well worth the small amount of time invested. Teachers generally reported that the retelling lasted about three minutes. A random sampling of the tapes confirmed

that fact. They also reported that their analyses took about five minutes for each retelling.

In summation, it was noted in the third chapter of this document that three assumptions were critical to this investigation. The first was that teachers would not give their students direct instruction in how to retell a story. The researcher saw no evidence to question this assumption, and the teachers confirmed that they gave students no such instruction. The second assumption was that teachers would make use of the diagnostic information they gathered to improve their instruction. Even though they were not told that this was a concern of the investigation, all six subjects restructured some portion of their teaching in order to use what they were learning from the procedure. The evidence in support of this assumption was abundant in both the informal conversations and the structured interviews, and it was observed by the researcher during each one of his visits to the teachers' classrooms. There can be no doubt that the teachers used the information they gained. Therefore, the qualitative data collected during this investigation tend to confirm the third assumption: the superior performance of the experimental group on the posttest of Experiment 3 appears to be attributable to the use of the retelling assessment procedure and the changes in instructional strategies that occurred as a result of that procedure.

CHAPTER FIVE
SUMMARY, IMPLICATIONS, CONCLUSIONS, AND RECOMMENDATIONS

Introduction

In recent years, a number of authors have suggested that oral retellings of the stories that children read can be used to diagnose reading comprehension. This investigation was designed to explore that suggestion and to determine the effects of the procedure on students and teachers. This chapter reviews the investigation and discusses its implications. Specifically, the chapter has three parts: (a) a summary of the investigation which includes the original research question and the four hypotheses that grew from it, a summary of the theoretical foundation of the study, and a description of the four studies conducted during the investigation and their results; (b) a discussion of the implications of this research and the conclusions that can be drawn from it; and (c) recommendations for the use of the technique and for future research.

Summary of the Investigation

The Research Question and the Four Hypotheses

This investigation focused on the following research question: Can the analysis of students' oral retellings of basal stories be used by classroom teachers as an effective

informal technique for diagnosing reading comprehension? This question arose from the suggestions of several authors (Clark, 1982; Irwin & Mitchell, 1983; Marshall, 1983; Morrow, 1984, 1985a, 1985b, 1985c) who believe that retellings can be used in the diagnosis and remediation of reading comprehension.

Four hypotheses were developed to provide the means for answering this research question. The first hypothesis stated that there would be no statistically significant correlation between students' standardized test scores and scores assigned to them through use of the oral retelling technique. This was intended as a test of the validity of the procedure used throughout this project. The second hypothesis stated that there would be no statistically significant correlation among the scores assigned to 10 different retellings of a single story by five trained users of the Carroll Oral Retelling Checklist (CORC).

The third hypothesis concerned the impact on students resulting from participation in the retelling activities and from the use of the procedure by their teachers. It stated that there would be no statistically significant difference between the final retellings of students who had been repeatedly exposed to the oral retelling technique and retellings of the same story by students who had not been exposed to the technique.

The final hypothesis focused on the impact that using the procedure might have on teachers' diagnostic insights. It stated that there would be no statistically significant difference between the diagnostic assessments developed by teachers who had used the oral retelling technique and by those who had not, as judged by a panel of experts in reading.

Testing these hypotheses involved a five-step process. First, an instrument had to be developed to use with the retelling procedure. Second, the validity and reliability of the instrument had to be established. Third, several teachers had to be trained in the use of the instrument. Fourth, the impact on students of the use of the technique by their teachers had to be assessed. Finally, the impact of the technique on teachers' diagnostic insights had to be assessed.

The research question and hypotheses noted above were the focal points of this investigation. Later in this chapter they are addressed in light of the findings described in Chapter Four and an answer to the question is offered. However, without a theoretical foundation for the investigation, an understanding of the implications of those findings would be impossible. The following section, therefore, briefly summarizes that foundation.

The Theoretical Foundation of the Investigation

As was noted above, the first step in this investigation was the development of an instrument for use in assessing oral retellings. The suggestion by Marshall (1983) that story schema research provides a highly efficient framework for such an instrument seemed promising. Thus, the first stage in designing the instrument was to conduct a review of the pertinent literature. However, a schema for stories is only one manifestation of the broader area of theory concerning the nature of schemata themselves. Therefore, the review began with the seminal work of Bartlett (1932).

Bartlett theorized that all of an individual's experiences are organized in memory in complex cognitive structures that he called "schemata." He believed that they are active constructs that give order to new stimuli by comparing them to prior experiences of a similar nature. He also believed that schemata evolve by changing to fit new data as the need occurs. Subsequent research has tended to confirm Bartlett's (1932) theories.

Recent investigators have expanded upon Bartlett's work (Rumelhart, 1980; Rumelhart & Ortony, 1977) and applied it to the reading process (Adams & Collins, 1979; Anderson & Pearson, 1984). Rumelhart and Ortony (1977) described schemata as constructs that have variables, that can be embedded within each other, that vary in their levels of

abstraction, and that represent knowledge. Rumelhart (1980) later added that they are active and that they are recognition devices that evaluate new data. Adams and Collins (1979) rejected the notion that the component levels of processing had to be hierarchical and argued instead that schemata must culminate in a central omniscient processor. Anderson and Pearson (1984) argued that schemata play a crucial role in the making of inferences, in allocating attention to text, and in selecting and organizing data during recall.

One particular schema, a schema for stories, has been widely investigated and the results of those investigations have tended to confirm the theories outlined above. Rumelhart (1975) presented a series of rewrite rules that describe the structure of most simple stories. His research provided evidence that the rules can be used to predict recall with a high degree of accuracy and tended to support the claims of schema theorists.

Mandler and Johnson (1977) provided substantial evidence in support of the crucial role played by one's story schema during comprehension and recall. They demonstrated that the patterns of recall followed the categories predicted by story structures and that well developed schemata in children and adults function in essentially the same manner. Mandler (1978) further demonstrated that readers use their story schema to

reorganize story elements during recall that were encountered in an inappropriate sequence during reading.

Kintsch (1977) reasoned that if schemata play such an important role in comprehension, then comprehension should suffer in the absence of an appropriate schema. He tested his hypothesis by using stories from another culture, and discovered that capable readers had great difficulty understanding and summarizing the material.

Stein and Glenn (1979) also examined the critical nature of a schema for stories. They replicated many of the findings of Mandler and Johnson (1977) and Mandler (1978) and expanded upon them. They demonstrated two additional points. First, contrary to previous findings, children are aware of the internal responses of characters in stories. Second, children do understand the causal relationships that link events in a story.

While all of the research cited above suggested that both children and adults have a schema for stories that functions in essentially the same manner, there were some important differences noted. Adults recalled more material than the older children who recalled more than the younger children, and each group did so with a slightly different saliency of categories (Mandler, 1978; Mandler & Johnson, 1977; Stein & Glenn, 1979). Furthermore, Hoover (1982) and Stein and Glenn (1982) have demonstrated that certain children may not be aware of some story categories nor of

the relationships that exist among story elements. These findings strongly suggest that the acquisition and use of a schema for stories are developmental in nature and that instruction in story structure may be of benefit to some children. Research into the effects of such instruction has tended to confirm its utility in improving comprehension (Bowman, 1980; Bowman & Gambrell, 1981; Cunningham & Foster, 1978; Fitzgerald & Spiegel, 1983; Singer & Donlan, 1982).

The literature reviewed above offers strong supporting evidence for the importance of schemata in the comprehension process and for the powerful effects of a schema for stories in understanding and recalling narratives. Furthermore, children's use of story schema and perhaps the acquisition of schemata seem to be developmental in nature. Also, instruction in the elements of a story appears to help enrich students' reading comprehension. Therefore, based on this evidence, Marshall's (1983) suggestion to use a story structure approach to the assessment of oral recalls would seem to be an excellent one and was adopted for use in this investigation.

The use of oral recalls, or retellings, in this investigation is based on work dating back to Courtis (1914), H. A. Brown (1914), and Starch (1915) and on the recent work of Clark (1982), Marshall (1983), and Irwin and Mitchell (1983, 1984). All of these authors presented techniques employing retellings in the assessment of

comprehension, but none of them presented empirical evidence in support of the claims that were made. Only Zimiles and Kuhns (1976) and Morrow (1984, 1985a, 1985b, 1985c) provided evidence of the effects on children of retelling stories. Both Morrow and Zimiles and Kuhns showed that children who do retellings immediately after hearing a story recall more over greater periods of time than children who do not retell them. Furthermore, Morrow showed that children who are prompted during a retelling with questions that focus on the structural elements of stories ultimately perform better retellings even in the absence of the prompts. Finally, Smith and Jackson (1985) reported success with a diagnostic and placement procedure that utilized written retellings of a specific selection by college students.

In summation, children apparently possess a schema for stories that exerts a powerful influence on their ability to comprehend narratives. Schema appears to be less well developed or less accessible in some children, especially poor readers. Therefore, a story-schema-theoretic approach to diagnosis and instruction would seem to be both justified for the teacher and beneficial for the student. Finally, the retelling of stories appears to be a convenient and efficient vehicle through which the adequacy of children's story schema can be assessed. Hence, this investigation combined the two positions in an attempt to discover a means

by which teachers can informally diagnose the quality of their students' reading comprehension.

Procedures and Findings

Four separate studies were designed to address the hypotheses noted above. The first sought to establish the validity of the CORC by determining the degree of correlation that exists between it and a widely used standardized test. Thirty fifth-graders were asked to read and retell a single story. Their retellings were assessed using the CORC and the results were correlated with the students' reading comprehension scores obtained on the Intermediate 1 level of the Metropolitan Achievement Test (Prescott, Balow, Hogan, & Farr, 1978). Results indicated that a significant direct relationship exists between the two instruments. The strength of the association varied from .57 using all data points, to .78 when three apparently invalid points were eliminated. Both correlations were highly significant ($p < .01$). Therefore, the first hypothesis was rejected.

The second investigation established the level of inter-rater reliability that can be expected when several users of the CORC assess the same retellings. Ten tape-recorded retellings of a single story were given to five educators who had been trained in the use of the technique. They assessed the quality of comprehension revealed in each retelling using the CORC. Correlations

were computed among all possible combinations of raters. The lowest correlation obtained was .86 and the highest .98 with 70% of the obtained correlations being above .90. All correlations were significant ($p < .01$), indicating that a significant degree of inter-rater reliabilitiy exists during use of the CORC. Therefore, the second hypothesis was also rejected.

The first two studies of this investigation showed the CORC to be a valid and reliable means of obtaining an informal diagnostic assessment of children's reading comprehension. The third and fourth studies were intended to explore the impact that repeated use of the procedure would have on students and teachers.

In the third study, the oral retelling assessment procedure was used in six Chapter I basic skills classes with a total of 30 fourth- and fifth-grade students. Teachers used the retelling procedure on a rotating basis with their students so that each student performed either five or six retellings over a 12-week period. The teachers were encouraged to use the information obtained during the retellings in any way they believed to be appropriate. The researcher held informal conversations with the teachers, conducted informal interviews, and observed their classes. Teachers reported that the procedure was giving them some highly usable information and all of them restructured their approach to teaching comprehension to take advantage of that

information. A posttest compared the retellings of the 30 students in the experimental group with those of 30 students in a control group who had been given practice in telling stories and in using the tape recorder. All 60 retellings were assessed using the CORC. An analysis of variance showed that the experimental group was rated significantly higher on the posttest than the control group ($F = 5.11$, $df = 1/58$, $p < .05$). This led to the rejection of the third hypothesis.

The final study was designed to assess the impact of repeated use of the technique on the diagnostic insights of teachers in the experimental group. Teachers in both the experimental and control groups were asked to listen to a single retelling of a story, to write in narrative form their professional assessments of the quality of the comprehension revealed in the retelling, and to prescribe a remedial treatment if appropriate. Three experts in reading independently assessed the twelve diagnoses and selected those that they believed were clearly better than the rest. Three z-tests of independent proportions were used to analyze the decisions of the experts. One of the tests proved to be significant in the predicted direction, favoring the assessments of the experimental group ($z = 2.286$, $p < .05$). The other two, however, were nonsignificant ($z = .597$ and $z = 0$). Hence, the fourth hypothesis could not be rejected.

Implications

The review of the literature in Chapter Two placed the use of story structure and retellings in their theoretical framework. Based on prior research, it was predicted that a story-schema approach would provide an efficient means for classifying students' comprehension of narratives. Retellings have been successfully used in reading research for a long time and recent authors have suggested that the retelling technique may prove to be an effective method for diagnosing comprehension. This investigation has provided evidence in support of both predictions.

The first study in this investigation indicated that the oral retelling assessment procedure described in Chapter Three provides a fairly accurate accounting of students' comprehension, especially when a student is not verbally inhibited. Results showed that a strong correlation exists between students' MAT reading comprehension scores and their retelling scores as determined by the CORC. Therefore, the CORC can be said to possess construct validity. That is, the CORC appears to be a valid instrument for the measurement of students' reading comprehension of narrative material.

In addition to construct validity, the CORC can be said to possess content validity. The retelling procedure can be used during a normal reading lesson with typical basal stories. As a result, it gives teachers an assessment of

their students' ability to comprehend the very material upon which instruction is based. Not all other types of informal diagnosis can claim both kinds of validity. For example, commercially prepared Informal Reading Inventories may accurately measure comprehension but they use preselected passages that may not be relevant daily reading tasks confronting students. The fact that the CORC seems to possess both construct and content validity is certainly to its advantage. However, unless it is also a reliable instrument its effectiveness would be limited.

The results of Experiment 2 indicate that a high degree of inter-rater reliability exists when several users of the CORC independently assess the same retellings. This finding was particularly robust across all 10 of the correlations. In fact, the lowest obtained correlation was highly significant ($r = .861$, $p < .01$). Combined with the results of Experiment 1, this suggests that the CORC is a valid and reliable instrument when employed as described in Chapter Three and can be used successfully by teachers to assess the quality of their students' reading comprehension. Furthermore, these two studies provide a foundation for the interpretation of the results of Experiment 3.

One of the principles of diagnosis cited by Spache (1976) is that all diagnoses should be tied to remediation. A valid and reliable instrument may be meritorious in itself but it is of little worth unless it can be used to plan a

remedial program. The third experiment conducted during this investigation provided both quantitative and qualitative data which confirm the effectiveness of the oral retelling assessment procedure as a diagnostic/prescriptive device. For ease of interpretation, the implications of this experiment are addressed under three general categories: the change that occurred in teachers' instructional practices as a result of their use of the CORC, the effects of that change on students, and the implications of these results with respect to the theory and research reviewed in Chapter Two.

Within a few weeks of the beginning of the third experiment, all six teachers in the treatment group altered their approach to the teaching of reading comprehension. This change was not limited to those students who were participating in the experiment; it extended to all of the children in each teacher's class. Some teachers adopted a story structure approach as an organizer for their discussion of the content of basal stories. Others used a strategy which employed probes constructed to reflect story-grammar categories. The probes were used to examine students' recall of the content of their basal selections. In all cases, teachers attributed the change in their instructional practices to the use of the oral retelling procedure.

One teacher mentioned that her students frequently were unable to explain the internal motivations of the characters in their stories, or tried to do so by explaining how they would react in a similar situation. These comments would seem to be in line with findings by Mandler and Johnson (1978) and Stein and Glenn (1979) concerning the saliency of the internal-response category. However, the teacher added that discussing stories in terms of a problem and a goal helped her students focus on internal motivations and improved their comprehension.

Another teacher made an interesting comment on the nature of the comprehension questions included in the basal series used in her school system. She believed that students could answer most or all of the questions correctly and still miss the point of the story and not understand the sequence and the rationale behind the events of the story. She asserted that the retelling procedure revealed more about the quality and depth of understanding achieved by her students than the basal questions did. She reasoned that if the six participating students showed little overall comprehension, then the rest of her class must be in a similar situation. Therefore, she changed her comprehension instruction by adopting a story structure approach as an organizer for prereading activities and subsequent discussions. She later indicated that she was satisfied

with the new strategy and that it seemed to help her students focus on content while reading.

The above anecdotes illustrate that the oral retelling procedure produced dividends for teachers on a global basis. They used their new insights to redesign their instruction to meet the needs of their students. Additional comments by teachers indicated that the CORC revealed specific weaknesses in the comprehension of several of the students participating in this experiment. Each teacher mentioned at least one instance, and often more, in which information gained while employing the CORC was used to design some specific remediation. For example, one student who had particular difficulty recalling events in a correct sequence was given additional practice with that skill. Several teachers mentioned that they had students who regularly failed to include details of the setting of their stories and frequently could not fill in that information during subsequent questioning. That weakness was alleviated to some degree, according to the teachers, but only after a substantial amount of questioning and instruction.

Based on the statistical results of Experiment 3, it appears that the use of the CORC for diagnostic and prescriptive purposes as indicated above has an important effect on the comprehension of students. Control-group students had an mean CORC score of 19.6 on the posttest. The experimental group had a mean of 23.5 on the CORC,

almost 20% better than the controls. As reported earlier, this difference is significant, indicating that the experimental students performed much better retellings during the posttest than can be explained by chance.

An explanation for this superior performance is easily culled from the comments made by teachers throughout the investigation. The unanimous opinion of the teachers in the experimental group was that their students lacked an adequate sense of the structure of stories. This made it more difficult for students to decide what was important to remember from a story and resulted in disjointed retellings that often focused on providing a large amount of details, without demonstrating a thorough understanding of the story.

This explanation clearly reflects the results of the research reported in Chapter Two. A number of authors (e.g., Hoover, 1982; Mandler & Johnson, 1977; Stein & Glenn, 1979, 1982) have demonstrated that the acquisition and use of a schema for stories is developmental in nature. Furthermore, the subjects of this investigation were below average fourth- and fifth-grade readers who have been shown by Canney and Winograd (1979) as tending toward viewing reading as a word-calling and not a meaning-obtaining activity. In this experiment, directed instruction aimed at improving students' sense of story structure and focusing their attention on meaning apparently led to their superior posttest performance. According to the teachers, that

instruction was initiated due to information obtained through the use of the CORC.

The statistical results of this experiment and the consensus of the opinions stated by teachers about their changes in instructional strategies can be summarized under three major points. First, through using the oral retelling assessment procedure the teachers became convinced that their students lacked a firm sense of the expected structure of stories. Second, that realization led them to change their instructional strategies for their entire class as well as for those individuals who were participating in the study. Third, the teachers believed that their students were performing better retellings at the end of the experiment because the new teaching strategies helped to improve their students' understanding of the structure of stories and thereby helped to improve reading comprehension.

The results of Experiments 1 and 3 tend to support this line of reasoning. Experiment 1 suggested that the CORC is a valid instrument for measuring students' comprehension of narratives. Therefore, the results of the posttest of Experiment 3 strongly suggest that the superior performance of the treatment group can be attributed to better comprehension. Furthermore, this offers confirmation of the position taken by Fitzgerald and Spiegel (1983) that instruction in the structure of stories can significantly

improve the ability of below average fourth-grade readers to comprehend stories.

The final concern of this investigation was the impact that repeated use of the oral retelling assessment procedure might have on the diagnostic insights of the teachers involved in the experimental group. Experiment 4 was designed to determine whether or not those teachers became better diagnosticians as a result of their participation. The qualitative results of Experiment 3 clearly showed that use of the procedure led to changes in instructional practices and helped to improve the comprehension of participating students. However, that experiment did little to determine whether use of the CORC resulted in an improved ability to diagnose comprehension weaknesses or provided a convenient framework for organizing teachers' thinking and using their professional diagnostic competence.

Unfortunately, the results of Experiment 4 do not resolve the issue. The lack of significance on the posttest and the failure to reject the fourth hypothesis may tend to suggest that the experimental group did not improve in the ability to diagnose comprehension through listening to a retelling of a story. However, offering that stance as a conclusion of this investigation would be inappropiate and would seem on the surface to be a direct contradiction of the findings of Experiment 3. At least two interpretations

of the data are possible and the following discussion explains both.

The first interpretation, as was implied above, is that all teachers participating in this study were already competent diagnosticians. The use of the CORC in the experimental classrooms, therefore, helped those teachers realize some things they had not noticed in the past, but did not affect their already high level of competence. It is true that the experimental-group teachers had indicated during conversations with the researcher that they believed it took two to three weeks to learn to use the procedure correctly. However, it may be that the time was needed to make them feel comfortable with the procedure but not to make their diagnoses any better; their first diagnosis may have been as valid as their last. Therefore, the lack of any demonstrable difference on the posttest may only indicate that teachers such as those in the two groups of this study already possess the ability to diagnose comprehension through listening to a retelling. The CORC, however, may provide them with an effective vehicle for using their competence to the benefit of their students.

An alternative interpretation is that Experiment 4 did not provide an adequate means through which teachers' diagnostic abilities could be measured. For example, the use of written assessments in this study may not have been an efficient means for determining the impact of the

retelling procedure on teachers' diagnostic abilities. Furthermore, the small sample size made it very difficult to obtain statistical significance, especially in light of the fact that one subject in the experimental group apparently did not take the posttest seriously and wrote a four-sentence diagnostic and remedial plan.

Another and perhaps more likely explanation for the inadequacy of the posttest used in Experiment 4 may be that the teachers and the experts did not agree among themselves on either the quality of the comprehension exhibited by the child in the posttest or on what constitutes an effective diagnosis. Vinsonhaler, Weinshank, Wagner, and Polin (1983) produced some disturbing evidence that reading professionals do not tend to agree with each other in their diagnoses or in their plans for remediation. That appeared to be the case in this study. One of the members of the panel of experts told the researcher that no comprehension weaknesses were readily apparent in the posttest retelling. Another said that numerous comprehension weaknesses could be identified. This basic lack of agreement among the experts may have contributed to the inconsistent statistical results.

Regardless of the interpretation placed on the data from this last experiment, the results of the three previous studies strongly suggest that the CORC can be an effective instrument in the diagnosis and remediation of students'

comprehension of narratives. The specific conclusions that can be drawn from those studies and some recommendations for future research are addressed in the next section of this chapter.

Conclusions and Recommendations

Conclusions

The quantitative and qualitative evidence gathered during this investigation seems to support eleven basic conclusions about the use of retellings in diagnosis and remediation, and about the critical role played by a schema for stories in several facets of this investigation. These conclusions, however, are valid only within the context of the limitations and delimitations of this investigation as described in Chapter One.

First, an oral retelling of a story as described in this document represents a fairly accurate picture of that student's reading comprehension.

Second, the previous conclusion is based on evidence that the Carroll Oral Retelling Checklist used throughout this investigation possesses both construct and content validity. The former is supported by the strong correlation between scores obtained through using the CORC and the standardized reading scores of the same students obtained from a widely recognized standardized test of reading comprehension. The latter arises from the fact that the CORC is designed to be used with the same basal stories that

are the typical instructional materials during normal reading lessons.

Third, the validity of the instrument is at its highest when students are at ease with the examiner and so feel free to report all that they remember from the story they have read. The validity is at its lowest point when students feel verbally inhibited and do not make a full report.

Fourth, the CORC is a highly reliable instrument. This study shows that educators who have received a minimal amount of training in the use of the CORC will produce a pattern of scores that is highly consistent when they are asked to evaluate the same retellings.

Fifth, there is some evidence to suggest that the process data section of the CORC can help direct teachers' attention to critical factors that affect reading comprehension. However, that evidence should be viewed as tentative pending future research.

Sixth, the story structure approach utilized in the product data section of the CORC provided a highly efficient means for classifying and evaluating students' reading comprehension. Retellings were quickly and easily analyzed using the categories of story grammars after only a small amount of training. The patterns of responses that emerged provided teachers with usable information and a clear picture of the quality of their students' understanding of the materials used during reading instruction.

Seventh, the evidence gathered during this investigation provides strong support for the developmental nature of a child's schema for stories. According to the earliest assessments and comments of the teachers who used the CORC in the experimental group, their students lacked an adequate sense of the structure of stories. This confirms findings of previous investigators (Hoover, 1982; Mandler, 1978; Mandler & Johnson, 1977; Stein & Glenn, 1979, 1982) that some students lack a fully developed schema for stories.

Eighth, the superior performance of the students in the treatment group on the posttest of Experiment 3 appears to be attributable to the development of their story schema due to the direct instruction they received in story structure. This confirms the findings of previous authors that such instruction can improve comprehension (Bowman, 1980; Bowman & Gambrell, 1981; Cunningham & Foster, 1978; Fitzgerald & Spiegel, 1983).

Ninth, the use of the CORC enabled teachers to realize that their students needed instruction in story-schematic categories and provided them with the information required to plan that instruction. Throughout this investigation, the CORC proved to be an effective instrument for diagnosing comprehension and for providing a framework for remedial instruction.

Tenth, the retelling assessment procedure and the CORC can be used on a frequent basis, during normal reading lessons, using typical basal materials, and with a minimum amount of effort. Consistent use of the technique requires little time and provides a continuing source of immediately relevant diagnostic information which can be used almost instantaneously to remediate weaknesses as they become apparent. Teachers in this study were unanimous in their belief that the procedure required very little time to use properly after an acclimation period of about two to three weeks. Furthermore, they agreed that the time spent was more than repaid by the dividends gained through accurate and usable diagnostic information.

Finally, the preponderance of evidence collected in this investigation supports the critical link between the process of comprehending narratives and story-schema theory. Several of the conclusions drawn above relate to individual applications of that theory. However, the evidence suggests that reading professionals should view the theoretical structure of stories as a construct that transcends all facets of the teaching of reading and of learning to read and to comprehend. It provides teachers with a framework for classifying retelling protocols, diagnosing comprehension weaknesses, and planning remedial instruction. Furthermore, when fully developed, it provides students with a set of expectations that guide and enhance comprehension.

In effect, this construct and its application in the oral retelling assessment procedure used in this investigation provide a direct link between theory and practice to the benefit of both teachers and students.

Recommendations

The recommendations that result from this investigation fall into two basic categories: those that concern pedagogical issues and the use of the CORC, and those that concern the course of future research into this topic.

Oral retellings should be adopted as a normal part of reading instruction. The evidence provided by Zimiles and Kuhns (1976) and Morrow (1984, 1985a, 1985b, 1985c) suggests that retelling stories improves subsequent recall, and this investigation indicates that retellings can be used to diagnose and remediate comprehension. The Carroll Oral Retelling Checklist that appears in Appendix A gives structure to the analysis of retellings and some direction to remedial efforts.

Young students, especially below average readers, apparently can benefit from some direct teaching in the structure of stories. The students in the experimental group of this investigation all received some form of instruction of this nature and with very significant results. That instruction appears to be effective when it is preceded by an accurate diagnosis of the status of students' schema for stories.

The oral retelling assessment procedures described in this document can be used with the CORC to obtain that diagnosis. However, the CORC is intended as only one form of diagnosis. It is, in fact, an informal technique and should be used to supplement and complement other diagnostic procedures.

With respect to future research, at least six recommendations can be made. First, in order to define more precisely the extent to which these findings are generalizable, the procedure and the CORC should be tested on other populations at different levels of proficiency. Second, an analysis of the scoring patterns on the checklists used by the teachers throughout Experiment 3 may reveal whether or not students' recalls followed patterns predicted by previous story-schema research. This could add another dimension to the validity of the CORC. Third, research should be conducted to determine the level of intra-rater reliability of the CORC. Such a study would provide additional evidence as to the reliability of the CORC when used by the same individual on the same retellings over a period of time.

Experiment 4 of this investigation should be redesigned to provide more definitive information on the impact of using the technique on teachers' diagnostic insights. One approach might be to analyze teachers' assessments to

determine if a significant difference exists in the quantity and quality of their references to the elements of a story.

The final two research areas to be mentioned here were suggested by teachers in the experimental group. The first suggestion centered on students' ability to make a metacognitive decision about their readiness to perform a retelling. Students could be instructed to read and study a story until they feel ready to retell it successfully by including all of the critical elements of the story in a coherent manner. Students' success with this task may serve as an indication of their awareness of their own comprehension successes and failures.

Finally, it is possible that the CORC can be used as a postreading activity by students to assess their own retellings. This activity may serve as a comprehension check and may contribute to an improvement in students' sense of the structure of stories by forcing them to analyze the products of their own comprehension along story-schematic lines.

Summary

The data reported in this document lead to two basic conclusions. First, this investigation supplies empirical evidence in support of the crucial role played by story schema in comprehension diagnosis and remediation. The categories of story schema provide an efficient means for classifying and assessing students' oral recalls, and the

information gained can be used to plan an effective remedial program. The evidence further indicates that instruction in the structure of stories results in improved comprehension of narratives, especially for children who are below average in reading.

Second, the evidence also supports the use of oral retellings in the classroom. Using retellings as described in Chapter Three is an effective means of accessing and diagnosing reading comprehension. Furthermore, the Carroll Oral Retelling Checklist is a valid and reliable instrument which can be used successfully in the analysis of oral recalls. It is more useful than an Informal Reading Inventory, less time consuming than a standardized test, and can be used with stories typically encountered during reading instruction.

These conclusions bring this paper back to the original research question: Can the analysis of students' oral retellings of basal stories be used by classroom teachers as an effective informal technique for diagnosing reading comprehension? In this study, the use of oral retellings effectively represented students' comprehension of stories, proved to be a highly reliable diagnostic technique, served as an efficient means for diagnosing comprehension weaknesses, provided the information necessary to plan a remedial course of action, and resulted in better comprehension for students in the experimental group.

Therefore, based on the evidence gathered during this investigation, an affirmative answer to the question appears to be justified.

APPENDIX A
CARROLL ORAL RETELLING CHECKLIST

Student Number_____ Date_____
School _____

Directions. Appoint one student to be the story teller of the day. Ask all students in reading group to read an entire story silently. After silent reading is complete have story teller retell the story as if the teacher and the rest of the reading group had never heard it before. At the end of the retelling ask the story teller the following question: "Is there anything else that you would like to add, or any comment that you would like to make?" As the retelling progresses, or immediately after it is completed, score the following items.

Product Data

Did the retelling include:	All	Much	Some	Little	None
Setting (eg. time, place, etc.)	_____	_____	_____	_____	_____
Main Characters	_____	_____	_____	_____	_____
Central Problem	_____	_____	_____	_____	_____
Goal	_____	_____	_____	_____	_____
Significant Events	_____	_____	_____	_____	_____
Resolution	_____	_____	_____	_____	_____
Summary Statement	_____	_____	_____	_____	_____
Theme or Moral	_____	_____	_____	_____	_____

Process Data

Did the retelling follow an appropriate sequence?

Did the retelling include any intrusions, and were they appropriate?

Did the retelling include any evidence of the student's metacognitive ability?

Did the retelling include any inferences, and were they accurate?

Considering the demands of the story, how adequate was the student's vocabulary?

Considering the demands of the story, how adequate was the student's background knowledge?

List areas of weakness: List areas of strength:

APPENDIX B
QUESTIONNAIRE FOR STRUCTURED INTERVIEWS

(1) Now that you have been using the retelling procedure for some time, do you have any overall impressions?

(2) Do your students understand what they have to do based on the directions on the form?

(3) Have you gained any valuable information from using the product-data section of the CORC?

(4) Have you gained any valuable information from using the process-data section of the CORC?

(5) In general, how have you used the data you have gained?

(6) Would you please think about each of the students who are participating in the experiment and comment on whether or not the CORC has helped you diagnose his or her comprehension weaknesses. If it has helped, how?

(7) Have you learned or realized anything about your students through using the CORC that you did not know before? If so, explain.

(8) In your opinion, is the time needed to use the CORC well spent?

(9) How long does it usually take your students to retell a story?

(10) How long does it usually take you to analyze a retelling?

(11) Do you intend to continue using the CORC after this study is completed? Why or why not?
If yes go to #12; if no go to #15.

(12) Would you use the CORC with all of your students or just some of your students? Explain.
If all, go to #15; if some go to #13.

(13) With what kinds of students are you most likely to use the CORC?

(14) With what kinds of students are you least likely to use the CORC?

(15) Would a retelling checklist covering other kinds of reading materials (non-narrative) be of value or interest to you?

REFERENCES

Adams, M. J., & Collins, A. (1979). A schema-theoretic view of reading. In R. O. Freedle (Ed.), New directions in discourse processing, Vol. II. Advances in discourse processing (pp. 1-22). Norwood, NJ: Ablex.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In P. D. Pearson (Ed.), Handbook of reading research (pp. 255-292). New York: Longman.

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. American Educational Research Journal, 14, 367-382.

Baker, L. (1979). Do I understand or do I not understand: That is the question (Reading Education Report No. 10). Urbana: University of Illinois, Center for the Study of Reading. (ERIC Document Reproduction Service No. ED 174 948)

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson (Ed.), Handbook of reading research (pp. 353-394). New York: Longman.

Bartlett, F. C. (1932). Remembering. Cambridge: Cambridge Univeristy Press.

Black, J. B., & Wilensky, R. (1979). An evaluation of story grammars. Cognitive Science, 3, 213-230.

Bobrow, D. G., & Norman, D. A. (1975). Some principles of memory schemata. In D. G. Bobrow & A. M. Collins (Eds.), Representation and understanding: Studies in cognitive science. (pp. 131-150). New York: Academic Press

Bowman, M. A. (1980). The effect of story structure questioning upon the comprehension and metacognitive awareness of sixth grade students. Dissertation Abstracts International, 42, 1052A. (University Microfilms No. 81-19,760)

Bowman, M. A., & Gambrell, L. (1981, April). The effects of story structure questioning upon reading comprehension. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.

Bransford, J. D., Stein, B. S., Shelton, T. S., & Owings, R. A. (1981). Cognition and adaptation: The importance of learning to learn. In J. Harvey (Ed.), Cognition, social behavior and the environment (pp. 93-110). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1980). Metacognitive development in reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 453-482). Hillsdale, NJ: Erlbaum.

Brown, H. A. (1914). The measurement of efficiency in reading. Elementary School Teacher, 14, 477-490.

Canney, G., & Winograd, P. (1979). Schemata for reading and reading comprehension performance (Tech. Rep. No. 120). Urbana: University of Illinois, Center for the Study of Reading. (ERIC Reproduction Service No. ED 169 520)

Clark, C. H. (1982). Assessing free recall. The Reading Teacher, 35, 434-439.

Clark, C. H. (1984, May). The Frecall method for evaluating free recall. Paper presented at the annual meeting of the International Reading Association, Atlanta, GA.

Clymer, T., Venezky, R. L., Johnson, D. D., & Pearson, P. D. (Eds). (1982). The Ginn Reading Program. Lexington, MA: Ginn.

Collins, A., Brown, J. S., & Larkin, K. M. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 385-410). Hillsdale, NJ: Erlbaum.

Courtis, S. A. (1914). Standard tests in English. Elementary School Teacher, 14, 374-392.

Courtis, S. A. (1917). Courtis standard research tests in silent reading. Detroit: Courtis.

Cunningham, J. W., & Foster, E. O. (1978). The ivory tower connection: A case study. Reading Teacher, 31, 365-370.

Davis, F. B. (1968). Research in comprehension in reading. Reading Research Quarterly, 4, 499-545.

Davis, F. B. (1972). Psychometric research on comprehension in reading. Reading Research Quarterly, 7, 628-678.

Dreher, M. J., & Singer, H. (1980). Story grammar instruction unnecessary for intermediate grade students. Reading Teacher, 34, 261-268.

Eller, W., & Hester, K. B. (1980). Penny's good fortune. In W. Eller & K. B. Hester (Eds.), The Laidlaw reading program: Level 10. Thundering giants (pp. 166-177). River Forest, IL: Laidlaw.

Fitzgerald, J., & Spiegel, D. L. (1983). Enhancing children's comprehension through instruction in narrative structure. Journal of Reading Behavior, 15 (2), 1-17.

Fry, E. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. Journal of Reading, 21, 242-252.

Glenn, C. G. (1977). Memory for multiple episodic stories: A developmental study. Paper presented at the annual meeting of the Psychometric Society, Washington, DC.

Goetz, E. T. (1979). Inferring from text: Some factors influencing which inferences will be made. Discourse Processes, 2, 179-195.

Goodman, K. S., & Goodman, Y. M. (1977). Learning about psycholinguistic processes by analyzing oral reading. Harvard Educational Review, 47, 317-333.

Goodman, Y. M. (1982). Retellings of literature and the comprehension process. Theory into Practice, 21, 301-307.

Goodman, Y. M., & Burke, C. L. (1970). Reading miscue inventory manual procedure for diagnosis and evaluation. New York: Macmillan.

Grant, E. (1980). Sybil's midnight ride. In W. Eller & K. B. Hester (Eds.), The Laidlaw reading program: Level 12. Patterns (pp. 164-169). River Forest, IL: Laidlaw.

Grant, J. (1984, May). Deciding what's important for a reader to remember. Paper presented at the annual meeting of the International Reading Association, Atlanta, GA.

Graves, M. F., Cooke, C. L., & LaBerge, M. J. (1983). Effects of previewing difficult short stories on low ability junior high school students' comprehension, recall, and attitudes. Reading Research Quarterly, 18, 262-276.

Harris, T. L., & Hodges, R. E. (Eds.). (1981). A dictionary of reading and related terms. Newark, DE: International Reading Association.

Hoover, N. L. (1982). Young children's detection of incoherence in narratives. In J. A. Niles & L. A. Harris (Eds.), New inquiries in reading research and instruction (269-273). Rochester, NY: National Reading Conference.

Irwin, P. A., & Mitchell, J. N. (1983). A procedure for assessing the richness of retellings. Journal of Reading, 26, 391-396.

Irwin, P. A., & Mitchell, J. N. (1984, May). Qualitative rating of retellings. Paper presented at the annual meeting of the International Reading Association, Atlanta, GA.

Johnston, P. H. (1983). Reading comprehension assessment: A cognitive basis. Newark, DE: International Reading Association.

Johnston, P. H. (1984). Assessment in reading. In P. D. Pearson (Ed.), Handbook of reading research (pp. 147-182). New York: Longman.

Kelly, E. J. (1916). The Kansas silent reading tests. Journal of Educational Psychology, 7, 63-80.

Kintsch, W. (1977). On comprehending stories. In M. Just & P. A. Carpenter (Eds.), Process in comprehension (pp. 33-62). Hillsdale, NJ: Erlbaum.

Kintsch, W., & Greene, E. (1978). The role of culture-specific schemata in the comprehension of stories. Discourse Processes, 1, 1-13.

Kintsch, W., & van Dijk, T. A. (1975). Recalling and summarizing stories. Languages, 40, 98-116.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Leys, M. (1984, May). Using retellings in the classroom: Four scoring systems. Paper presented at the annual meeting of the International Reading Association, Atlanta, GA.

Mandler, J. M. (1978). A code in the node: The use of a story schema in retrieval. Discourse Processes, 1, 14-35.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 9, 111-151.

Marshall, N. (1980, May). Informal assessment of reading: An application of reading research. Paper presented at the annual meeting of the International Reading Association, St. Louis, MO.

Marshall, N. (1983). Using story grammar to assess reading comprehension. Reading Teacher, 36, 616-620.

Marshall, N. (1984a). Discourse analysis as a guide for informal assessment of comprehension. In J. Flood (Ed.), Promoting reading comprehension (pp. 79-96). Newark, DE: International Reading Association.

Marshall, N. (1984b, May). Story grammar as a basis for evluating a retelling. Paper presented at the Annual Conference of the International Reading Association, Atlanta, GA.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision (pp. 211-277). New York: McGraw-Hill.

Mitchell, J. N. (1984, May). What retellingts reveal about readers. Paper presented at the Annual Conference of the International Reading Association, Atlanta, GA.

Mitchell, J. N., Clark, C., Grant, J., Irwin, P., Marshall, N., Mason, J., & Leys, M. (1984, May). Building classroom reading comprehension with free recalls or "retellings." Preconvention Institute presented at the annual meeting of the International Reading Association, Atlanta, GA.

Morrow, L. M. (1984). Effects of story retelling on young children's comprehension and sense of story structure. In J. A. Niles & L. A. Harris (Eds.), Changing perspectives on research in reading/language processing and instruction (pp. 95-100). Rochester, NY: National Reading Conference.

Morrow, L. M. (1985a). Reading and retelling stories: Strategies for emergent readers. Reading Teacher, 38, 870-875.

Morrow, L. M. (1985b). Retelling stories: A strategy for improving young children's comprehension, concept of story structure and oral language complexity. Elementary School Journal, 85, 647-661.

Morrow, L. M. (1985c, May). Story retelling: A diagnostic approach for evaluating story structure, language and comprehension. Paper presented at the annual meeting of the International Reading Association, New Orleans, LA.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

Pikulski, J. J., & Shanahan, T. (1982). Approaches to the informal evaluation of reading. Newark, DE: International Reading Association.

Prescott, G. A., Balow, I, H., Hogan, T. P., & Farr, R. G. (1978). Metropolitan achievement test. Intermediate level. New York: The Psychological Corporation.

Rosenshine, B. V. (1980). Skill hierarchies in reading comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 535-554). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. M. Collins (Eds.), Representation and understanding (pp. 211-236). New York: Academic Press.

Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading comprehension (pp. 265-303). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 33-58). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of Knowledge (pp. 99-136). Hillsdale, NJ: Erlbaum.

Schank, R. C., & Abelson, R. P. (1975). Scripts, plans and knowledge. In Advance papers of the fourth international joint conference on artificial intelligence. Tbilisi, Georgia, USSR.

Shell, F. (1977). Brush with white death. In M. W. Meyer, P. Travers, J. Hatch, & V. Heller (Eds.), Reading skill builder. Level 5. Part 1 (pp. 88-94). Pleasantville, NY: Reader's Digest.

Singer, H., & Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation for comprehension of complex short stories. Reading Research Quarterly, 17, 166-186.

Slosson, R. L. (1963). Slosson Oral Reading Test. Aurora, NY: Slosson Educational Publications.

Smith, S. P., & Jackson, J. H. (1985). Assessing reading/ learning with written retellings. Journal of Reading, 28, 622-631.

Spache, G. D. (1976). Diagnosing and correcting reading disabilities. Boston: Allyn and Bacon.

Spearitt, D. (1972). Identification of subskills of reading comprehension by maximum likelihood factor analysis. Reading Research Quarterly, 8, 92-111.

Starch, D. (1915). The measurement of efficiency in reading. Journal of Educational Psychology, 6, 1-24.

Stein, N. L. (1979). How children understand stories: A developmental analysis. In L. G. Katz (Ed.), Current topics in early childhood education, Vol II (pp. 261-290). Norwood, NJ: Ablex.

Stein, N. L., & Glenn, C. G. (1978). The role of temporal organization in story comprehension (Tech. Rep. No. 72). Urbana, IL: University of Illinois, Center for the Study of Reading.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), New directions in discourse processing, Vol. II: Advances in discourse processing (pp. 53-120). Norwood, NJ: Ablex.

Stein, N. L., & Glenn, C. G. (1982). Children's concept of time: The development of a story schema. In U. J. Friedmen (Ed.), The developmental psychology of time. New York: Academic Press.

Thorndike, R. L. (1973). Reading as reasoning. Reading Research Quarterly, 9, 137-147.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 9, 77-110.

Tierney, R. J., & Cunningham, J. W. (1984). Research on teaching reading comprehension. In P. D. Pearson (Ed.), Handbook of Reading Research (pp. 609-655). New York: Longman.

Turner, A., & Greene, E. (1977). The construction of a propositional text base (Tech. Rep. No. 63). Boulder, CO: University of Colorado Press.

Van Dijk, T. A. (1977). Text and context: Explorations in semantics and pragmatics of discourse. London: Longman.

Vinsonhaler, J. S., Weinshank, A. B., Wagner, C. C., & Polin, R. M. (1983). Diagnosing children with educational problems: Characteristics of reading and learning disabilities specialists, and classroom teachers. Reading Research Quarterly, 18, 134-164.

Zimiles, H., & Kuhns, M. (1976). A developmental study of the retention of narrative material, final report. New York: Bank Street College. (ERIC Document Reproduction Service No. ED 160 978)

BIOGRAPHICAL SKETCH

Robert G. Carroll was born in Scranton, Pennsylvania, on November 18, 1949. He attended St. Patrick's Grade and High School until he graduated in 1967. He has received a Bachelor of Arts in English literature (1971), a Master of Science in English education (1974), and a Master of Science in reading (1976) from the University of Scranton.

He began his teaching career at Bishop Hoban High School in Wilkes Barre, Pa., in 1971. Three years later he was hired by the Scranton Public School District where he taught English, developmental reading and remedial reading. In 1982 he moved to Gainesville where he began work on his Ph. D. at the University of Florida.

He is married to Gennette Gailey of Gainesville, Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

H. Thompson Fillmer, Chairman
Professor of Instruction and Curriculum

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Nora L. Hoover
Assistant Professor of Instruction and Curriculum

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stephen F. Olejnik
Associate Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Suzanne M. Kinzer
Associate Professor of Instruction and Curriculum

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1985

David C. Smith
Dean, College of Education

Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08285 364 8